Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *University of Michigan*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Cross-Modality Validation of Subjective Scales for Loudness, Vibration, and Electric Shock: S. S. STEVENS 201
Tactile Vibration: Dynamics of Sensory Intensity: S. S. STEVENS 210
Effect on Extinction of Restricting Information in Verbal Conditioning: O. E. ROGERS, W. B. WEBB, AND T. J. GALLAGHER 219
Frustrative Factors in Selective Learning With Reward and Nonreward as Discriminanda: A. AMSEL AND D. L. PROUTY 224
Sweetness and Saltiness of Compound Solutions of Sucrose and NaCl as a Function of Concentration of Solutes: J. G. BEEBE-CENTER, M. S. ROGERS, W. H. ATKINSON, AND D. N. O'CONNELL 231
The Gelb Effect: E. C. STEWART 235
Intersensory Comparisons of Temporal Judgments: S. GOLDSTONE, W. K. BOARDMAN, AND W. T. LHAMON 243
Performance in Eyelid Conditioning as a Function of UCS Duration: W. N. RUNQUIST AND K. W. SPENCE 249
The Relation of Achievement in Problem Solving to Rate and Kind of Hypotheses Produced: R. J. REINE 253
Warm-up in Retention as a Function of Degree of Verbal Learning: J. E. DINNER AND C. P. DUNCAN 257
Meaningfulness and Articulation of Stimulus and Response in Paired-Associate Learning and Stimulus Recall: R. G. HUNT 262
Stimulus Generalization After Training on Three Stimuli: A Test of the Summation Hypothesis: H. I. KALISH AND N. GUTTMAN 268

American Psychological Association

Vol. 57, No. 4 April 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 57, NO. 4 APRIL, 1959

CROSS-MODALITY VALIDATION OF SUBJECTIVE SCALES FOR LOUDNESS, VIBRATION, AND ELECTRIC SHOCK [1]

S. S. STEVENS

Harvard University

These experiments attempt to compare the dynamic characteristics of three different sense modalities. Once the ratio scales for subjective intensity have been determined on several perceptual continua, an interesting question arises. Can these scales be used to predict what people will do when asked to make direct matches between the apparent strengths of sensation on pairs of these continua? More specifically, since the subjective magnitude of a prothetic continuum grows approximately as a power function of stimulus intensity (S. S. Stevens, 1957), can the relations among the exponents of different power functions be verified by cross-modality matches at various stimulus levels?

This problem was investigated with three modalities: loudness, vibration, and electric shock. These three were chosen because the intensities of their respective stimuli can be conveniently controlled by electrical networks, and, more important, because the three exponents involved are very different in size. These three continua also have the advantage that the typical *O* is so unfamiliar with the physical measures of the respective stimuli that the "stimulus error" can probably be ruled out as exceedingly unlikely.

Extensive research in several laboratories has shown that loudness grows approximately as the .6 power of the sound pressure (or the .3 power of the sound energy) (S. S. Stevens, 1955). Although studies of the other two modalities have been less thorough, it seems quite well established that the exponents for vibration and electric shock (both applied to the fingers) are decidedly larger than the exponent for loudness. By the method of magnitude estimation, under which *O* tries to assign numbers to various stimulus intensities in a manner that reflects the apparent intensities of the sensations produced, it has been shown that the apparent intensity of a mechanical vibration (60 cps) applied to the finger tip grows approximately as the .95 power of the vibration amplitude (S. S. Stevens, 1959). By a similar method,

[1] Research supported by Grant G-2668 from the National Science Foundation and Contract Nonr-1866(15) with the Office of Naval Research (Project NR 142-201, Report PNR-216). Reproduction for any purpose of the U. S. Government is permitted.

the apparent intensity of a 60-cycle current applied to the fingers has been found to grow roughly as the fourth power of the current (Stevens, Carton, & Shickman, 1958). Approximate and uncertain as may be their exact values, there is little question that the exponents for loudness, vibration, and electric shock differ significantly in size. They are so different, in fact, that only a modest degree of experimental precision is needed to verify that the differences exist.

If the numerical estimations that people make when asked to judge their sensations reflect something fundamental concerning the dynamic operating characteristics of the transducers involved, then the exponents of the functions obtained should predict the approximate slopes of the curves that result from cross-modality comparisons. Furthermore, if the ratio scales for two subjective magnitudes are power functions, cross-modality comparisons should produce a function that can be plotted as a straight line in log-log coordinates. In general, if, given an appropriate choice of units, two continua are governed by the equations

$$\psi_1 = \phi_1^m$$

$$\psi_2 = \phi_2^n$$

and if the psychological values ψ_1 and ψ_2 are equated at various levels, then it follows that the stimulus values ϕ_1 and ϕ_2 should stand in the relation

$$\log \phi_1 = \frac{n}{m} \log \phi_2$$

In other words, the expected slope of the line relating ϕ_1 and ϕ_2 in a log-log plot is given by the ratio of the exponents. If one or the other of the two sensory continua departs significantly from a power function, the log-log plot of ϕ_1 versus ϕ_2 should produce a curve rather than a straight line, but the amount of the curvature should be predicted by the forms of the two magnitude functions. The ability of cross-modality matches to verify these predictions constitutes a necessary condition (but not a sufficient condition) for the validity of the several subjective scales.

Apparatus and Procedure

Provision was made to stimulate the observer (*O*) with sound, vibration, and electric current by means of the apparatus schematized in Fig. 1. Pairs of stimuli could be presented either simultaneously or successively. The level of one stimulus of a pair was fixed and the level of the other stimulus was adjusted by *O*. Each kind of stimulus served sometimes as the fixed and sometimes as the variable stimulus. The levels of the stimuli were measured on the decibel scale of a Ballantine vacuum-tube voltmeter, supplemented by calibrated attenuators.

The sound used was a band of frequencies obtained by passing a white noise through a UTC-4C filter with cut-off frequencies set at 100 and 500 cps. The choice of this particular noise was arbitrary, but preliminary exploration seemed to indicate that

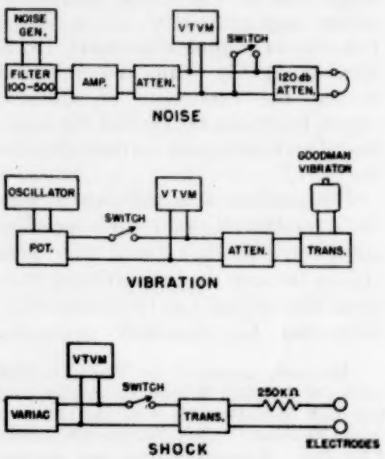

Fig. 1. Diagram of the apparatus.

it was relatively satisfactory to work with. Delivered through PDR-8 earphones mounted in sponge Neoprene cushions, MX-41/AR, it becomes a broad band of noise having a maximum intensity in the vicinity of 500 cps. The *O* adjusted the level of the noise by means of an attenuator (1-db steps) having a range of 120 db.

The vibration was delivered to the middle finger of the left hand by means of a Goodmans Vibration Generator (Model V47) kindly loaned by F. A. Geldard. The vibrator was driven at 60 cps. The *O* rested his finger on a plastic button (¼ in. in diameter) driven by the vibrator. He adjusted the amplitude of the vibration by means of a control consisting of two 2000-ohm potentiometers ganged and cascaded.

Shock was produced by 60-cycle current passed through the fingers of the left hand. The *O* placed his first and third fingers in liquid electrodes consisting of two glass jars containing salt solution. These jars were connected through a transformer to a Variac, which provided the current control. A resistor of 250,000 ohms was placed in series with the electrodes. In the experiments on shock, care was taken to avoid using *O*s who had cuts or abrasions on the fingers.

In the process of adjusting the stimulus level, *O* used the method of "bracketing," turning the level alternately too high and too low in order to "zero in" on equality. He was not allowed to see the dial on the stimulus control. Some 30 different *O*s participated in one or another part of these experiments. They were all staff members or students, and were used in the experiments more or less at random, depending on their availability. Experimental sessions lasted about 10 min. each.

Magnitude Scales

As already indicated, the ratio scale for subjective loudness has been very thoroughly investigated, especially for pure tones (Robinson, 1957; S. S. Stevens, 1955). Except at levels lower than about 40 db above threshold, the loudness of noise grows much like the loudness of a 1000-cycle tone. By the method of magnitude estimation the exponent of the power function for noise has turned out in several experiments to lie in the vicinity of .6 (J. C. Stevens, 1958; Stevens and Tulving, 1957). Two unpublished studies have given a similar value. For the purposes of comparisons to be made below, it seems reasonable to take this value .6 as the approximate exponent governing

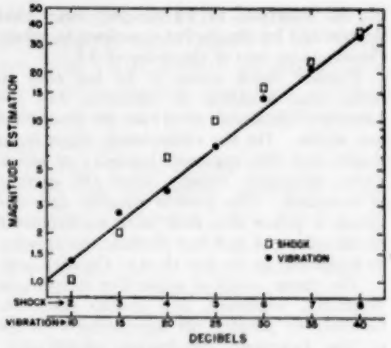

FIG. 2. Magnitude estimations of the apparent intensity of mechanical vibration on the finger tip (circles), and electric current through the fingers (squares). Vibration follows a power function; shock follows a function four or five times steeper than vibration but departs from a power function.

the growth of loudness as a function of sound pressure.

A subjective scale for vibration was determined with 10 *O*s, each of whom placed the middle finger of his left hand on the button of the Goodmans vibrator in such a way that the axis of the finger was approximately parallel to the direction of the vibration. The *O* was first stimulated with a moderate amplitude and told to call it 10. He was then asked to assign numbers proportional to the apparent intensity of the various stimulus levels, which were presented in a more or less random order. Each stimulus level was presented twice (duration about 1 sec.). The geometric means of the 20 estimates are plotted as circles in Fig. 2. Over the range of stimuli from 10 to 40 db above threshold the numerical judgments approximate a power function of vibration amplitude, with a slope (exponent) of about .96.

A similar procedure was used with 10 *O*s to obtain magnitude estimations of the apparent intensity of electric current (60 cps) delivered to the first and third fingers of the left hand. The geometric means of the estimates are represented by the squares in Fig. 2. The squares determine a power function much less exactly than do the circles. As a matter of fact, the squares show a downward concavity not unlike that suggested by the data reported earlier (S. S. Stevens, Carton, & Shickman, 1958). There is also a tendency for the over-all slope

of the function to be steeper than that determined by the earlier experiments, where the exponent was of the order of 3.5.

Electric shock seems to be less easy to judge than loudness or vibration, and the functions obtained seem to be somewhat less stable. On the other hand, there is no doubt that the apparent intensity of shock grows extremely rapidly when the current is increased. The growth roughly approximates a power function with an exponent of the order of 4 or 5, but the function appears to be steeper at the low than at the high end.

The three scales of subjective magnitude (loudness, vibration, and shock) can now be used to predict the approximate forms of the functions (in log-log coordinates) that should be obtained when *O*s try to equate apparent intensity in one of these modalities to apparent intensity in another. Vibration versus loudness should yield a straight line with a slope of about .6, since that is the ratio of the two exponents. Shock versus loudness should give an over-all slope of the order of .13, but since the magnitude function for shock is somewhat curved, the matching function should also exhibit curvature. Shock versus vibration should give an over-all slope of about .2, again with evidence of curvature.

The following experiments test these predictions.

Vibration vs. Loudness

The matching of the apparent intensities of sound and vibration was carried out in two complementary experiments. In one experiment the level of the sound was adjusted to match the vibration; in the other the level of the vibration was adjusted to match the sound. Both sound and vibration were presented simultaneously. Each of 10 *O*s made two adjustments at each level in each experiment.

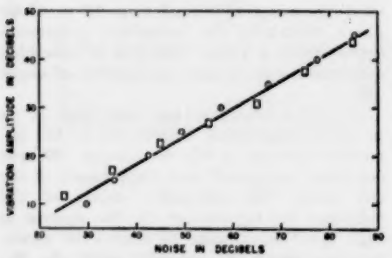

Fig. 3. Equal-sensation function for vibration and loudness. Squares mean vibration was adjusted to match loudness; circles mean loudness was adjusted to match vibration.

The results are shown in Fig. 3. The circles represent the means of the decibel levels to which the sound was adjusted, and the squares represent the means of the decibel levels to which the vibration was adjusted. The coordinate scales are in decibels relative to the approximate thresholds of the two kinds of stimuli.

The interesting point to note is that the slope of the line in Fig. 3 is .6, which is almost exactly the slope that is called for by the ratio of the exponents of the two magnitude functions. It is also apparent that the relation is essentially linear, which is consistent with the fact that, over the ranges of the stimuli involved, both loudness and vibration are governed by power functions.

The departure of some of the points from the straight line in Fig. 3 is in large measure due to the interesting fact that, depending on which stimulus is adjusted, the slope turns out to be slightly different. The situation is analogous to the two regression lines in a correlation plot. This "regression" or "centering tendency" is common, if not universal, in matching procedures, and it points up the necessity of a balanced design in which each stimulus is made to serve as both the standard and the variable (cf. S. S. Stevens, 1955). The size of the regression effect seems to depend on several factors, the most important perhaps being the difficulty of the judgment involved.

The matching of loudness and vibration turned out to be surprisingly easy. Some of the *O*s, who happened

to have served in loudness matching experiments, expressed the belief that matching loudness to vibration was in some ways easier than matching the loudnesses of two tones of widely different pitch or quality (cf. S. S. Stevens, 1956a). The consistency of the judgments seemed to bear this out.

Shock vs. Loudness

The matching of shock to loudness and loudness to shock was carried out with groups of 10 *O*s. Each *O* made two adjustments at each level. In order to minimize the adaptation that takes place when current is passed through the fingers, the shock and the sound were presented alternately (about 1-sec. durations). Even with this alternating presentation some adaptation was evident. For this reason, in Fig. 4 the averages of the first adjustments are plotted as unfilled symbols and the averages of all adjustments are plotted as filled symbols. The effects of adaptation are noticeable at the low current levels, but seem to be negligible at high current levels.

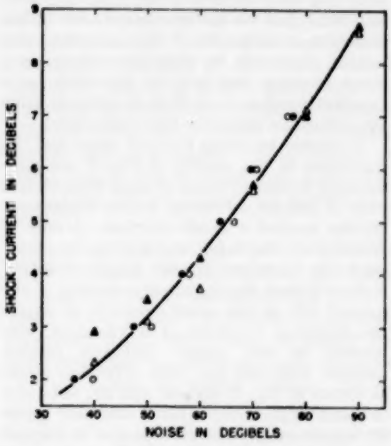

Fig. 4. Equal-sensation function for shock and loudness. Triangles mean shock was adjusted to match loudness; circles mean loudness was adjusted to match shock. Unfilled symbols represent means of first adjustments; filled symbols represent means of all adjustments. Some adaptation to shock is evident at low levels.

In Fig. 4 the circles represent the results of adjusting loudness; the triangles represent the results of adjusting shock. Here again there is evidence of a "regression" effect: the slope is a little steeper when the loudness is adjusted than it is when the shock is adjusted.

Two features of the data in Fig. 4 are noteworthy. The over-all slope is of the order of .13, which is about the slope predicted from the magnitude functions for loudness and shock. Even more interesting, perhaps, is the fact that the function in Fig. 4 is curved in a manner consistent with the fact that in a log-log plot the loudness function is straight and the shock function is concave downward (Fig. 2).

Shock vs. Vibration

Two pairs of counterbalanced experiments were run to match shock and vibration. In the first experiment, which, incidentally, was chronologically the first experiment in the present series, *O* held his finger on the vibrator in such a way that the axis of the middle finger was at right angles to the axis of the vibration. Each of 7 *O*s made one adjustment of the shock to match a vibration and one adjustment of the vibration to match a shock. The mean results are shown by the unfilled symbols in Fig. 5. Circles mean that the shock was adjusted and squares mean that the vibration was adjusted.

In an effort to make the matching judgment a little easier and to check on the curvature evident in the first experiments (which the writer had

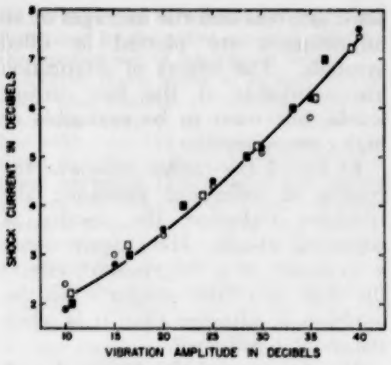

FIG. 5. Equal-sensation function for shock and vibration. Circles mean shock was adjusted; squares mean vibration was adjusted. Unfilled symbols represent an experiment in which the middle finger was at right angles to the direction of the vibration. Filled symbols represent an experiment in which the middle finger was held parallel to the direction of the vibration.

not at the time expected), in the second pair of experiments the axis of the finger was placed parallel to the axis of the vibration, i.e., the end of the finger met the vibrator button head on. Otherwise the procedure was the same as before. The results are shown by the filled symbols in Fig. 5, whose coordinates are in decibels relative to the approximate thresholds.

The different finger positions had little effect on the outcome, although what difference there is coincides with the very slight difference in the respective magnitude functions (see S. S. Stevens, 1959). The exponent of the power function for apparent intensity of vibration is a trifle smaller when the finger is at right angles, and the slightly lower slope determined by the unfilled points in Fig. 5 seems to confirm this fact. This is a very fine point, however, and it is far from certain that these differences would stand up under repetition.

More important to the present purpose is the fact that the over-all slope in Fig. 5 is of the order of .2, which closely approximates the predicted value. Furthermore, the curvature is in the direction predicted by the two magnitude functions in Fig. 2.

As in the other matching experiments, a slight "regression" effect is also evident in Fig. 5.

A word is in order regarding the problem of averaging the results from the different *O*s in the foregoing experiments. Since all *O*s do not have the same threshold, especially for electric shock, it was not always possible to use the same absolute range of stimuli for all *O*s. When this was not possible, a constant number of decibels was added to or subtracted from the absolute values used, so that approximately the same range of sensation levels (decibels re threshold) was obtained. The need for such corrections was, however, relatively infrequent. In general, since it is only the *form* of the equal-sensation functions that we are concerned with in the matching experiments, it is permissible, and indeed necessary, to shift the values on a given function one way or the other by a constant number of decibels in order to bring the individual functions into coincidence.

It should be noted that the slope and the curvature of the results in Fig. 5 are reassuringly similar to those of some unpublished data of Békésy, obtained in this laboratory. Békésy applied a small electrode (2 mm. in diameter to the finger and a large "neutral" electrode elsewhere on the hand. Each of 3 *O*s adjusted the apparent intensity of the current felt at the small electrode to match the apparent intensity of a vibration, also applied to the finger. Békésy's results, coupled with the fact that vibration grows as shown in Fig. 2, suggest that the sensation of shock applied via a small electrode grows by approximately the same law as the sensation produced with liquid electrodes into which the fingers are immersed.

DISCUSSION

The matching of the apparent intensities of sound, vibration, and shock has, in effect, closed an important circle in a

process of validation. Knowing the ratio scales of subjective magnitude we can apparently predict the results of matching experiments. Likewise, having constructed a ratio scale of subjective magnitude for one modality, and having matched that modality to a second modality, we can deduce the magnitude scale for the second modality. This latter procedure was essentially what was undertaken when, following the first experiment on shock versus vibration, an experiment was run to determine whether the scale for shock (Fig. 2) is (*a*) as steep and (*b*) as curved as the results in Fig. 5 suggest.

All in all, the outcome of these cross-modality validations is such that there seems to be little left to be said. The method of magnitude estimation, and the other ratio-scaling procedures, seem to tell us something about the dynamics of sensory processes. Despite the considerable variability that attaches to numerical estimates of sensory effects, it seems clear that the typical, intelligent person can give meaningful estimates of relative subjective magnitude. Over and over again the averages (geometric means or medians) of the estimates of groups of ten or so people have shown a high degree of stability whenever reasonable care has been taken to eliminate constraints and biases. What some of these biases result from has been discussed in a previous study (S. S. Stevens, 1956b).

As has been emphasized elsewhere (S. S. Stevens, 1958b) each of the ratio scaling methods (e.g., magnitude estimation) seems to have a certain amount of inherent bias, which can probably be corrected only by the use of one or more of the other methods. These biases are mainly "regression" effects, not unlike those exhibited in the matching experiments (Fig. 3–5). Thus the method of magnitude estimation tends in general to underestimate the slope of the magnitude scale relative to the estimate obtained by "production" procedures, in which *O* sets the stimulus to some prescribed ratio or magnitude (cf. Stevens & Poulton, 1956). On the other hand, since the predicted forms of the equal-sensation functions here described are all based on a common method (magnitude estimation), they should not be affected by a small constant bias in the method.

The question naturally arises whether matching is not by itself an adequate procedure, capable of answering whatever questions may arise regarding sensory dynamics. Cross-modality matching can reveal the *relative* forms of the dynamic growth functions in the different sense modalities, but, unfortunately, it cannot disclose the absolute form of any one of these functions. (This is what was meant by the statement above that the success of cross-modality matching constitutes a necessary but not a sufficient condition for accepting the validity of the several subjective scales.) Figure 3, for example, shows that the apparent intensity of vibration on the finger tip grows about 1.5 times as rapidly as loudness when the decibel levels of the stimuli are increased by a given amount, but, from this fact alone, the form of the sone scale could not be deduced unless the form of the vibration scale were already known. If the forms of the sensory scales are to be known, at some point, it seems, the form of at least one of the sensory scales must be determined by one or another process of direct judgment. The success of the foregoing matching experiments suggests that, in principle, the form of only one sensory scale need be determined in order to make it possible to fix, by matching, the forms of all the others. On the other hand, it is comforting to be able to verify the form of a scale by more than one method, as has been done here.

Whereas the matching of apparent magnitudes may be used to verify only the *relative* forms of two subjective scales, there is a technique by which it can be decided whether the two subjective scales are power functions. This procedure calls for the direct matching of apparent ratios, as exemplified by an adjustment of the ratio between two loudnesses to equal the apparent ratio between two brightnesses (for results

obtained by J. C. Stevens, see S. S. Stevens, 1957). In principle, if for all possible pairs of such ratio matchings the stimulus ratio between one pair (e.g., lights) equals the stimulus ratio between the other pair (e.g., sounds) raised to a power, it follows that the psychological magnitudes are power functions of their respective stimuli (S. S. Stevens, 1958b) This, in effect, is a direct test of the principle that equal stimulus ratios correspond to equal sensation ratios, which is another way of saying that the psychophysical law is a power function. Ratio matching can tell us whether two subjective scales are power functions and it can tell us the *ratio* of the two exponents, but it cannot tell us the the absolute values of the exponents.

Another interesting question is whether similarly good predictions regarding the outcome of matching experiments could have been made if Fechnerian scales for the several modalities had been constructed from measurements of just noticeable differences. The answer appears to be No. Newman (1933) has shown, for example, that JND scales are not capable of predicting the results of matching experiments (equal-loudness contours) even within a single sense modality (see also Stevens & Davis, 1938). Thus, the fact that ΔI is consistently smaller for 1000 cps than for 200 cps would suggest that the loudness function for 1000 cps is steeper than that for 200 cps. Actually, direct matching shows that the reverse is true.

On the three continua we are concerned with here, differential sensitivity is governed by the equation $\Delta I = kI + C$, which is Weber's law with the addition of a constant which behaves in some respects as a residual noise factor. The constant k has approximately the value .05 for loudness (in terms of sound pressure) (Miller, 1947), the value .05 for vibration (Békésy, 1930; Knudsen, 1928), and the value .02 for shock (as determined on one O in this laboratory by A. Jacobson and R. Wylie). From values such as these, it would not be possible to construct scales consistent with the results in Fig. 3, 4, and 5. For example, with the constant C neglected, these values of k suggest, if anything, that the slope of the line in Fig. 3 should be close to unity, not .6. Of course, since they were determined on different Os by different methods, these values of k may or may not be comparable. In the present state of psychophysics there seems to be little relation between differential sensitivity and the slopes of the various scales of subjective magnitude. The widespread feeling that some systematic relation *must* exist has thus far proved unfounded.

One of the uses of subjective ratio scales rests on their ability to make clear the dynamics of the processes served by the sensory transducers. In the three modalities investigated we have examples of transducers that appear to have three radically different operating characteristics. The slow growth of loudness (exponent less than one) suggests that the ear behaves as a "compressor" (S. S. Stevens, 1958a; Stevens, Carton, & Shickman, 1958). This compressor action probably helps to make it possible for the ear to respond to an enormous range of sound pressures—a range of millions to one. The apparent intensity of vibration on the finger tip grows almost linearly with vibration amplitude—as though the transducer were approximately linear. The effective range of vibration amplitudes to which the finger is sensitive is of the order of hundreds to one. (Incidentally, vibration on the arm does not follow a simple power law [S. S. Stevens, 1959].) The steep operating characteristic for electric shock suggests the action of an "expander" of some sort: doubling the current increases the sensation about tenfold. And correlated with this rapid expansion is a narrow operating range of stimuli—of the order of only tens to one.

The task of explaining these operating characteristics in terms of the action of the various transducers and their associated neural networks presents an interesting challenge. Needless to say, a full account is not yet possible. The task attempted in the foregoing experi-

ments was to determine the dynamic operating characteristics of the three modalities, not to advance the reasons for them. Before explanations are undertaken, it may be profitable to try to discover what it is that needs explaining.

Summary

The numerical ratio scales of subjective magnitude for three modalities (loudness, mechanical vibration on the finger tip, and electric shock to the fingers) were used to predict the form of the equal-sensation functions that should result from matching the apparent intensity in one modality to that in another, e.g., loudness adjusted to appear equal to a given vibration on the finger. Cross-modality matches were made between each modality and each of the other two. The stimulus of a given modality served one time as the standard and one time as the variable to be adjusted.

To a good approximation, the slopes and the curvatures of the three predicted functions were recovered experimentally. The ability of scales obtained by magnitude estimation to predict the form of the equal-sensation functions seems to attest the validity of subjective estimates, and to justify their use in studies directed at the dynamic characteristics of sensory systems.

References

Békésy, G. v. Uber das Fechnersche Gesetz und seine Bedeutung für die Theorie der akustischen Beobachtungsfehler und die Theorie des Hörens. *Ann. Physik*, 1930, **7**, 329–359.

Knudsen, V. O. "Hearing" with the sense of touch. *J. gen. Psychol.*, 1928, **1**, 320–352.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.*, 1947, **19**, 609–619

Newman, E. B. The validity of the just noticeable difference as a unit of psychological magnitude. *Trans. Kansas Acad. Sci.*, 1933, **36**, 172–175.

Robinson, D. W. The subjective loudness scale. *Acustica*, 1957, **7**, 217–233.

Stevens, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.

Stevens, J. C., & Tulving, E. Estimations of loudness by a group of untrained observers. *Amer. J. Psychol.*, 1957, **70**, 600–605.

Stevens, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.

Stevens, S. S. The calculation of the loudness of complex noise. *J. acoust. Soc. Amer.*, 1956, **28**, 807–832. (a)

Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25. (b)

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Stevens, S. S. Measurement and man. *Science*, 1958, **127**, 383–389. (a)

Stevens, S. S. Problems and methods in psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196. (b)

Stevens, S. S. Tactile vibration: Dynamics of sensory intensity. *J. exp. Psychol.*, 1959, **57**, 210–218.

Stevens, S. S., Carton, A. S. & Shickman, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, 1958, **56**, 328–334.

Stevens, S. S., & Davis, H. *Hearing: Its psychology and physiology*. New York: Wiley, 1938.

Stevens, S. S., & Poulton, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.

(Received May 26, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

TACTILE VIBRATION: DYNAMICS OF SENSORY INTENSITY [1]

S. S. STEVENS

Harvard University

The purpose was to determine, by the method of magnitude estimation, how the subjective intensity of a vibratory stimulus varies with the amplitude of a vibration applied to the finger and to the arm, and to compare the function obtained on the finger with two other scales: the category scale and the Fechnerian scale constructed from just noticeable differences (JND's). One question is whether vibration behaves as a prothetic continuum and exhibits the same features that characterize other continua of this class (Stevens, 1957b). A further question concerns the extent to which subjective estimates of apparent intensity are able to reveal the dynamic "operating characteristics" of sensory systems. If the sensory responses mediated by two different groups of receptors follow different growth functions, can the method of magnitude estimation disclose the fact, and can the relation between the two functions be verified by a procedure of direct matching? The validity of the ratio scales of sensory magnitude and their utility in solving problems related to the dynamics of sensory transducers depend on the answers to these questions.

The term "dynamic" is used here to designate the functional input-output relations of a sensory system, i.e., the growth of sensation as a function of stimulus intensity. It contrasts with the "static" relations that are ordinarily determined by the use of *O* as a null instrument in the mapping of equal-sensation contours. Although more attention has usually been paid to static relations (e.g., luminosity functions) it seems clear that in a complete account of the sensory process the other dimension—dynamics—must also be included.

[1] Research supported by Grant G-2668 from the National Science Foundation and Contract Nonr-1866(15) with the Office of Naval Research (Project NR142-201, Report PNR-215). Reproduction for any purpose of the U. S. Government is permitted.

VIBRATION ON THE FINGER

Ratio Scale

Three different experiments were run to determine a ratio scale for subjective vibration applied to the fingers. One experiment involved all the fingers and the two others involved only a single finger.

Experiment I.—The first experiment (conducted with the assistance of G. M. Shickman) was mainly exploratory. Each of 12 *O*s made magnitude estimations of the apparent vibration of a metal rod held between the thumb and the fingers. The rod was $\frac{1}{8}$ in. in diameter and was driven at a frequency of 120 cps by an electrodynamic driving unit (PAL).

Before each series of judgments *O* was presented with a moderate amplitude (25 db in Fig. 1) and told to call it 10. Then the various stimulus levels were presented twice each in "random" order and *O* assigned numbers proportional to the apparent intensity. Each presentation lasted about 1 sec.

The geometric means of the magnitude estimations are shown by the circles in Fig. 1. These values fall close to a straight line (log-log plot), which indicates that subjective vibration grows as a power function of stimulus amplitude. The exponent of the power function, given by the slope of the line, is about .83.

Experiment II.—In this experiment the method of stimulation was different. Instead of holding a vibrating rod, *O* rested his middle finger on a plastic button (¼ in. in diameter) whose direction of vibration was normal to the surface of the finger. The button was driven at 60 cps by a Goodmans Vibration Generator (Model V 47) kindly loaned for this purpose by F. A. Geldard. The *O* rested his arm on a cushion and maintained a light steady pressure on the button with his finger, which was held approximately at right angles to the direction of the vibration.

Except for this difference in mode of stimulation, the procedure was identical to that of Exp. I.

The geometric means for 15 *O*s are shown by the triangles in Fig. 1. The slope (.94) of the line through the triangles is steeper than that through the circles. Whether this difference is due to the manner or frequency of stimulation or to the difference in *O*s is not known for certain (only 5 *O*s served in both experiments). On the other hand since other experiments still in progress have shown that the slope of the function varies systematically with the frequency of the vibration applied to the middle finger, it seems probable that the difference in the slopes of the two upper lines in Fig. 1 is due to the frequency rather than to the direction of stimulation.

The abscissa of Fig. 1 designates the stimulus level in decibels relative to the approximate threshold of the vibration, as determined for three *O*s with the Goodmans vibrator. The weakest stimulus used was about 10 db above threshold, i.e., its amplitude was about three times the minimal detectable amplitude.

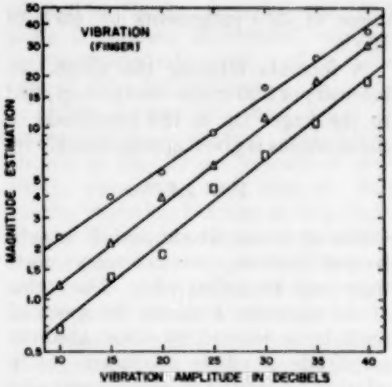

FIG. 1. Magnitude functions for apparent intensity of vibration. Circles: *O* held a vibrating rod between his fingers. Triangles: *O* rested his middle finger on a vibrating button, holding it at right angles to the direction of vibration. Squares: *O* rested his middle finger on the button, holding it parallel to the direction of vibration. Points are geometric means of the estimates. For greater clarity the positions of the curves have been separated vertically. The abscissa scale represents level above the approximate threshold.

Experiment III.—This experiment differed from Exp. II in only one respect. Instead of holding his middle finger at right angles to the direction of the vibration, *O* placed the tip of his finger on the button in such a way that the axis of the finger was roughly parallel to that of the vibration. In this position the vibration tends to feel a little stronger, but the slope of the function seems to be little affected. The exponent determined by the squares in Fig. 1 is about .96. Each square represents the geometric

mean of two judgments by each of 10 *O*s.

A formula relating the subjective intensity of a 60-cycle vibration applied to the finger tip to the amplitude of the stimulus is given approximately by

$$V = kA^{.95}$$

where A is amplitude and V is subjective intensity in arbitrary units that may be called *vibs*. The value of the constant k cannot be specified until it is decided at what absolute amplitude a vib is produced. Only relative amplitudes were measured in these experiments.

The exponent is, of course, the most interesting part of this equation. The exponent for 60-cycle vibration on the finger is a little more than half again as large as the exponent (.6) governing the loudness of a 1000-cycle tone (as a function of acoustic amplitude). On the other hand, since the loudness of very low tones grows more rapidly than the loudness of a 1000-cycle tone (Robinson & Dadson, 1956), it is probable that the growth of vibration and loudness would be more nearly the same if the stimulus were 60 cps in both cases.

It should be noted that there is some evidence that the method of magnitude estimation tends slightly to underestimate the size of the exponent of the power function, as compared to the estimations reached by the methods of ratio or magnitude production (e.g., fractionation). If this is true, it may be that the exponent that would emerge from a thorough and balanced program of research on 60-cycle vibration applied to the finger tip would not differ appreciably from 1.0. Unfortunately, the exponent for vibration has not had the theoretical and commercial importance that attaches to the exponent for loudness, and that has led to the scaling of loudness in several laboratories in at least four different countries (Robinson, 1957; Stevens, 1955).

Category Scale

Since a procedure that tries to get *O* to make judgments of equal-appearing intervals on a prothetic continuum always results in a function that is inconsistent with the results of magnitude estimation, a test was made to see whether this also holds for vibration.

Experiment IV.—The same stimuli represented by the triangles in Fig. 1 were used to determine a category scale for vibration. A stimulus of 10 db was presented and called 1 and then a stimulus of 40 db was presented and called 7. The *O* was instructed to judge the various levels on a scale from 1 to 7. Each of the 10 *O*s judged each level twice in a "random" order.

The mean category assignments are plotted as circles in Fig. 2. The abscissa scale is the subjective scale determined by the line through the triangles in Fig. 1. The data are plotted in this way in order to show that the category scale is nonlinearly related to the ratio scale of subjective magnitude. On prothetic continua,

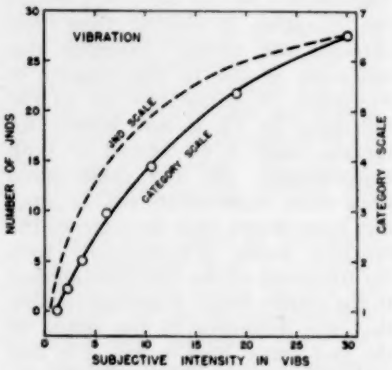

FIG. 2. Category scale and JND scale for vibration, plotted against the scale of subjective magnitude determined by the triangles in Fig. 1.

this nonlinearity appears to be an invariant feature of all methods that try to get *O* to partition a segment of the continuum into equal-appearing intervals. The curvature of the category scale can be altered to some extent by stimulus spacing, but the fact that it is always a decelerating or convex function of subjective magnitude (i.e., is concave downward) is presumably due to the fact that the impressiveness of differences varies from one end to the other of a prothetic continuum (see Stevens & Galanter, 1957). The convexity determined by the circles in Fig. 2 merely confirms what has previously been observed in category scales on some 10 other prothetic continua (Stevens, Carton, & Shickman, 1958; Stevens & Galanter, 1957; Stevens & Stone, 1959).

JND Scale

Difference limens for amplitude of vibration were not measured in these experiments, but measurements of differential sensitivity were recently reported by Geldard (1957). Interestingly enough these measurements can be approximated by a linear equation, $\Delta I = kI + C$ which is the same type of equation that seems to hold on a wide variety of prothetic continua. It is Weber's law modified by the addition of a constant. Approximately this same form of function was reported earlier by Knudsen (1928) and by Békésy (1930) for vibration applied to the finger tip. Both these experimenters found that the value of k was about .05.

On the assumption that an equation of this form would hold for differential sensitivity to vibration under the conditions of the present experiment, a JND scale was constructed (for Geldard's data) and plotted in Fig. 2. As usual, the JND scale is concave downward, relative to the magnitude scale, and is more curved than the category scale. The relations among the three types of scales (ratio, category, and JND, shown in Fig. 2) are typical of prothetic continua, and suggest that tactile vibration belongs to this class. Unlike the situation on a metathetic continuum, such as pitch (Stevens, 1954), on the prothetic continuum of vibration the subjective size of the JND grows rapidly larger as intensity is increased. It does not stay constant as Fechner had assumed.

Vibration on the Arm

The fact that the sensed magnitude of vibration applied to the fingers follows a power function adds interest to the finding that vibration applied to the arm produces a significantly different function. Two experiments were run to determine this fact.

Experiment V.—In this experiment the button of the Goodmans vibrator was applied to the under surface of the forearm, about 3 in. from the elbow. The *O* placed his arm across two supports separated by 2.5 in. The vibrator button was midway between the supports and about $\frac{1}{4}$ in. below their top surfaces. The fleshy part of the forearm then rested lightly on the button. Ten *O*s made magnitude estimations of the strength of the vibration (60 cps) relative to a standard (15 db in Fig. 3, called 10), which was presented at the beginning of each run, as in Exp. I–III.

The geometric means of the judgments are shown by the circles in Fig. 3. On the basis of casual observation, plus the results reported by Békésy (1958), the writer had expected that the growth of apparent intensity of vibration would be steeper

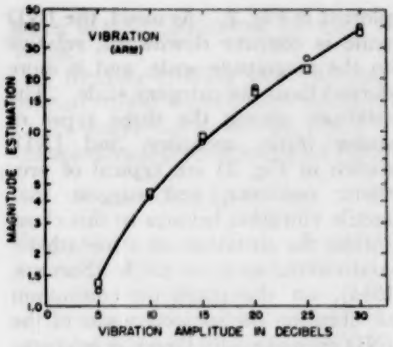

FIG. 3. Magnitude functions for apparent intensity of vibration applied to the arm. Squares and circles represent two different experiments. The abscissa scale represents level above the approximate threshold.

on the arm than on the finger, but the curvature shown in Fig. 3 was not anticipated. Distrust of it led to another experiment.

Experiment VI.—On the assumption that the closeness of the supports holding the arm may have interfered with the sensation at low stimulus levels, the supports were moved about 7 in. apart, and the experiment was repeated. Of the 10 *Os* used in this repeat experiment, 6 had served in Exp. V and 4 were new to the procedure.

The position of the arm rests seems not to have mattered, for the geometric means shown by the squares in Fig. 3 confirm the curvature found in Exp. V. The proximity of the squares and circles also gives some idea of the stability of these functions when replicated on groups of 10 *Os*. It is clear from Fig. 3 that over the low stimulus levels the perceived strength of a vibration grows more rapidly on the arm than on the finger. Over the higher ranges (15 to 30 db above threshold) the growth of sensation on the arm approximates a power function with a slope not very different from that obtained on the finger. Actually, at the top end of the scale the slope for the arm seems to be a little flatter than that for the finger.

It should be noted that, although the range of stimuli applied to the arm was less than that applied to the finger, the subjective ranges were almost precisely the same, as will be shown below. What happens to subjective magnitude outside the ranges here used is not yet known.

THE MATCHING OF APPARENT INTENSITY

Given the two different functions shown in Fig. 1 and 3, one for the finger and one for the arm, a crucial question arises. If the functions obtained by magnitude estimation are valid indications of the dynamics of sensory response, it should be

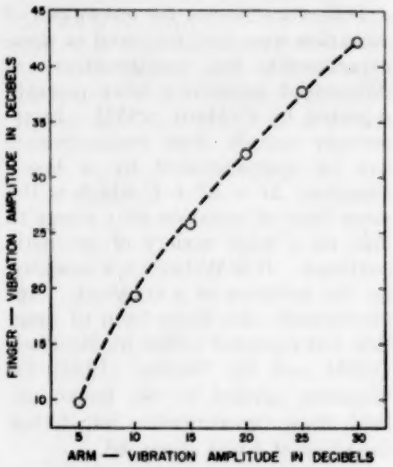

FIG. 4. Equal-sensation function for vibration on finger and arm. The dashed curve is the function predicted from the relation between the curve in Fig. 3 and the triangles in Fig. 1. Points are observed values. Coordinate scales are relative to the approximate thresholds.

possible to predict what people would do if asked to adjust a vibration on the finger to match the apparent intensity of a vibration on the arm.

From the two functions (triangles in Fig. 1 and curve in Fig. 3) the form of the equal-sensation function to be expected from a matching experiment was derived. This function is shown by the dashed line in Fig. 4. It was constructed by determining what decibel change on the finger is required to produce a change in magnitude estimation equal to that produced by a given decibel change on the arm. Thus a 5-db change on the arm (lower end of curve in Fig. 3) produces nearly a fourfold change in *O*'s estimate; and a fourfold change on the finger requires approximately 10 db. Hence, the bottom end of the dashed curve has a slope of approximately 2. Where the two magnitude functions have the same slope, the curve in Fig. 4 has a slope of 1. And at the top end the slope is a little less than 1.

This procedure determines the *form* of the dashed curve, not its vertical or horizontal position. Actually, it turns out that the absolute sensitivity of the arm is less than that of the finger. Thus the absolute threshold is about 15 db lower on the finger. And in order to match a strong stimulus on the arm, the stimulus on the finger requires an absolute amplitude that is about 10 db lower. These values agree fairly well with those reported by Békésy (1955). The position of the dashed curve in Fig. 4 was adjusted to reflect these facts.

The circles in Fig. 4 represent the means of 20 adjustments, two by each of 10 *O*s. The *O* rested his left arm on the Goodmans vibrator, as in Exp. VI, and simultaneously rested his middle finger on a button fastened to the PAL driving unit, the one used in Exp. I. The two vibrators were energized alternately for about 1 sec. at a time (60-cycle current), and *O* adjusted the stimulus applied to the finger to match the sensations produced at the arm. Most *O*s reported that the match was a relatively easy one to make. The matches can certainly be said to agree quite well with the predicted contour.

It should be mentioned that both vibrators were checked for linearity with the aid of a General Radio vibration pick-up. In addition, the similarity of the responses of the two vibrators under loading by the finger was checked by placing a finger alternately on the two vibrators and adjusting them to apparent equality. This matching with the finger (by 2 *O*s) gave a straight line with a slope of 1.0 in decibel coordinates.

It is interesting to note that vibration applied to parts of the body other than the arm may also produce a sensation that grows more rapidly than on the finger. Matching experiments have shown this to be true for the shoulder, for example (Békésy, 1955), and also for the toe (Périlhou & Piéron, 1942).

Discussion

As already indicated, the evidence in Fig. 4 provides an important confirmation of the general validity of people's numerical estimates of apparent intensity on a sensory continuum. Needless to say, this does not mean that everybody is capable of making valid estimates. Some people use numbers in rather strange ways (see Stevens, 1956). But the typical numerical estimates made by groups of *O*s chosen more or less at random from among the staff and students of this laboratory seem to be lawful and meaningful.

In the judgments of the relative strengths of the various vibrations

applied to the arm, the results were consistent whether these *Os* estimated the strengths numerically or matched them with a stimulus applied to the finger. This is at least encouraging, as is also the fact that similar consistency has been shown when magnitude functions in two different modalities have been checked by cross-modality matches (Stevens, 1957b; Stevens, 1959).

There is, however, an important difference between matching and magnitude estimation. By means of matching alone, it would never be possible to learn the form of the function relating sensory response to stimulus intensity. It appears that only magnitude estimation (combined with the other ratio-scaling methods [Stevens, 1958a]) can reveal the nature of this relation. We can go from the two magnitude functions (Fig. 1 and 3) to the curve in Fig. 4, but if given only the data in Fig. 4, we could not construct a magnitude function. Matching can verify the relative forms of magnitude functions, but it cannot produce these functions. This is not to say, however, that matching does not have an important place in psychophysics. It provides us with equal-loudness contours, luminosity functions, etc., but such functions are in a sense "static" functions. By themselves they tell us nothing about the "dynamics" of the sensory processes. A complete understanding of the sensory domain would seem to call for an exploration of both its static and its dynamic aspects.

The fact that, on most prothetic sensory continua, subjective magnitude approximates a power function of stimulus magnitude (Jones, 1958; Stevens, 1957b; Stevens, Carton, & Shickman, 1958; Stevens & Stone, 1959) makes it doubly interesting when a departure from this relation turns up. The power function, with its implication that equal sensation ratios correspond to equal stimulus ratios, is a first-order "psychophysical law," but significant departures from it are not infrequent. Interestingly enough, the most obvious instances involve departures in the same direction: the magnitude function becomes convex (concave downward) in log-log coordinates, much like the function for vibration on the arm. Examples: the loudness of low-frequency tones, the loudness of tones in ears that exhibit recruitment, and the brightness of lights viewed by a highly light-adapted eye. (These functions for the light-adapted eye have been determined in this laboratory by J. C. Stevens.) For other examples, see Stevens (1958b). The writer is aware of no established instance of a subjective magnitude function that is consistently concave upward in log-log coordinates, although upward concavity may sometimes occur locally (Stevens, 1957a), and also under certain types of multiple stimulation, such as combinations of tones widely spaced in frequency (Zwicker, Flottorp, & Stevens, 1957).

Is the fact that several sensory systems depart from the power law in the same general direction merely a coincidence, or is there some common mechanism involved? First, it should be noted that convex functions (in log-log coordinates) arise under two different circumstances. (*a*) A continuum that normally follows a power law may be made convex by fatigue, masking or disease. So-called recruitment in hearing is one example. The analogous phenomenon of recruitment in a light-adapted eye, and in an eye subject to glare, is another. (*b*) A few continua exhibit a convex magnitude function even in what appears to be their normal state. Vibration on the arm and the loudness of low-frequency tones are two examples. Whether the same causes underlie the convexity in both the "normal" and the "abnormal" case is an interesting question.

Recruitment in the ear has sometimes been held to be due to a deficiency of functional "neural elements," which may arise from masking, overstimulation or disease (see Stevens & Davis, 1938). Recruitment in the light-adapted eye may have a similar cause. The hypothesis is that a given deficiency in the available neural elements would result in a large elevation of the threshold without causing a proportionately large

change in the intensity required to produce a given superliminal effect. Needless to say, this is a vague hypothesis, for the precise neural mechanisms involved in these processes are not yet fully understood.

Do the sensory continua that normally exhibit convexity have something in common with an overstimulated ear or a light-adapted eye? Quite possibly. One suggestive fact is that the arm is less densely innervated than the finger. In a sense, therefore, the arm exhibits a "deficiency" in neural elements (Békésy, 1955). This fact is suggestive, but it may not be the whole story. (For more of the story on some of these neural mechanisms, see Békésy [1958].)

Whatever the mechanism may be, a convex function in log-log coordinates often accompanies an elevated threshold. This fact raises the question whether the two phenomena may be related in such a way that a correction for the threshold would remove the convexity. In other words, if we add a constant A_o to the measured value of the stimulus, will a power function of the form

$$V = k(A - A_o)^n$$

turn out to fit the data? The addition of A_o may be regarded as a device to alter the zero on the physical scale in order to make it coincide with the zero on the subjective scale. As shown in Fig. 5, the curved function for vibration on the arm (Fig. 3) becomes a near approach to rectilinearity if, from each of the amplitude values, we subtract a constant amplitude that is roughly equal to the threshold value. This outcome suggests that the foregoing equation may be a useful generalization of the psychophysical law relating the ratio scale of apparent magnitude to the ratio scale of stimulus magnitude.

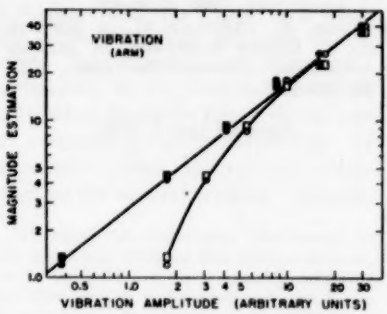

FIG. 5. The magnitude function for vibration on the arm becomes approximately straight in log-log coordinates when a constant value is subtracted from each stimulus value. The unfilled points are plotted as in Fig. 3. The filled points are the corrected values. They approximate a power function.

It is an interesting question whether an additive correction, applied to the stimulus scale, will eliminate the curvature in the log-log plots of some of the other instances in which there is an apparent departure from the power law. When the downward curvature occurs in the region near threshold, as it seems to do rather frequently, this simple correction may prove effective in bringing the empirical function into agreement with the power-function form. To the extent that this procedure proves effective, it will serve to augment the empirical evidence that the power function is the general, and perhaps the exclusive, form taken by the functional relation between stimulus and sensory magnitude.

SUMMARY

The ratio scale for the subjective intensity of vibration (60 cps) applied to the finger tip was shown to be a power function of vibration amplitude. Under the method of magnitude estimation the exponent of the power function turned out to be about .95. For vibrations (120 cps) tangential to the surface of the fingers the exponent was about .83.

Judgments of vibration on a 7-point category scale gave a function that is concave downward relative to the ratio scale obtained by magnitude estimation.

The scale of JND's is also concave downward, and is more curved than the category scale.

The growth of the subjective intensity of vibration (60 cps) applied to the arm does not follow a power law. In a log-log plot the function is concave downward. On the arm the threshold is higher than on the fingers,

but over the low stimulus levels the sensation on the arm grows more rapidly with vibration amplitude. The difference between the two subjective magnitude functions (arm and finger) was confirmed by having *Os* match the apparent intensity on the finger to that on the arm. The agreement suggests that the functions obtained by magnitude estimation are valid measures of the dynamic processes in sensory systems.

REFERENCES

BÉKÉSY, G. v. Ueber das Fechnersche Gesetz und seine Bedeutung für die Theorie der akustischen Beobachtungsfehler und die Theorie des Hörens. *Ann. Physik*, 1930, **7**, 329–359.

BÉKÉSY, G. v. Human skin perception of traveling waves similar to those on the cochlea. *J. acoust. Soc. Amer.*, 1955, **27**, 830–841.

BÉKÉSY, G. v. Funneling in the nervous system and its role in loudness and sensation intensity on the skin. *J. acoust. Soc. Amer.*, 1958, **30**, 399–412.

GELDARD, F. A. Adventures in tactile literacy. *Amer. Psychologist*, 1957, **12**, 115–124.

JONES, F. N. Scales of subjective intensity for odors of diverse chemical nature. *Amer. J. Psychol.*, 1958, **81**, 305–310.

KNUDSEN, V. O. "Hearing" with the sense of touch. *J. gen. Psychol.*, 1928, **1**, 320–352.

PÉRILHOU, P., & PIÉRON, H. Quelques caractéristiques des sensations vibratoires. *C. R. Soc. Biol.*, 1942, **136**, 448–449.

ROBINSON, D. W. The subjective loudness scale. *Acustica*, 1957, **7**, 217–233.

ROBINSON, D. W., & DADSON, R. S. A redetermination of the equal-loudness relations for pure tones. *Brit. J. appl. Physics*, 1956, **7**, 166–181.

STEVENS, S. S. Pitch discrimination, mels and Kock's contention. *J. acoust. Soc. Amer.*, 1954, **26**, 1075–1077.

STEVENS, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

STEVENS, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.*, 1957, **29**, 603–606. (a)

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181. (b)

STEVENS, S. S. Problems and methods in psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196. (a)

STEVENS, S. S. Some similarities between hearing and seeing. *Laryngoscope*, 1958, **48**, 508. (b)

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57**, 201–209.

STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, 1958, **56**, 328–334.

STEVENS, S. S., & DAVIS, H. *Hearing: Its psychology and physiology*. New York: Wiley, 1938.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

STEVENS, S. S., & STONE, G. Finger span: Ratio scale, category scale, and JND scale. *J. exp. Psychol.*, 1959, **57**, 91–95.

ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. Critical bandwidth in loudness summation. *J. acoust. Soc. Amer.*, 1957, **29**, 548–557.

(Received June 3, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

EFFECT ON EXTINCTION OF RESTRICTING INFORMATION IN VERBAL CONDITIONING [1]

OWEN E. ROGERS, WILSE B. WEBB,[2] AND THOMAS J. GALLAGHER [3]

U. S. Naval School of Aviation Medicine

Recently, Estes and Straughan (1954) and others (e.g., Grant, Hake, & Hornseth, 1951) have used the experimental procedure of "verbal conditioning" to test or develop a number of aspects of "statistical" learning theories. Basically, the procedure follows the original Humphreys (1939) experiment in which *S* is asked to predict on each trial whether a light will or will not come on. Over a number of trials the light does come on at some predetermined random ratio schedule. The proportion of positive guesses and the persistence of response in absence of reinforcement are usually the dependent variables.

Our interest in this procedure came from a different direction. We raised this question: Suppose on each trial *S* was given the additional alternative of neither "yea" nor "nay," and when this alternative was adopted he received no information about the probability of the occurrence of the light, what would be the effect on rate of responding and persistence of response? This contingency interested us for several reasons. Experimentally, it seemed to approximate rather closely many instrumental conditioning situations, e.g., the Skinner box in which the bar is presented to the rat and he obtains information about payoff only on the contingency of a response; no response yields no information. Perhaps more to the point, it represented an often-occurring paradigm of human behavior in which the individual has the option of trying a task or not but can achieve information about the task only by trying.

The same ratios of reinforcement used by Estes and Straughan (1954), 30%, 50%, and 85%, were used with two learning conditions. Half of the *S*s guessed on each trial and received partial reinforcement in random relation to these responses; the other half were permitted to guess or not guess on each trial and receive partial reinforcement only on the trials when they guessed. In addition to assessing the effects on rate of response (which was the concern of Estes and Straughan), resistance to extinction was also measured.

METHOD

Apparatus.—The *S*s were seated in individual booths. A telegraph key was on a platform extending 12 in. from the front of the booth and was 38 in. from the floor. The reinforcing light (GE-NE51 neon lamp) was 3½ in. above this platform in the front wall of the booth.

The entire room was air conditioned and all booths were observable from a soundproof control room. As many as 19 *S*s could be run simultaneously.

[1] Opinions or conclusions contained in this paper are those of the authors and are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department. The first author, OER, now deceased, suggested the experimental situation; WBW designed the experiment and is responsible for the analysis and writing; TJG collected the data.

[2] Now at the University of Florida.

[3] Now with the US Naval Aircrew Equipment Laboratory, Johnsville, Pa.

The presentation of the signal for each trial (a buzzer) was automatically controlled. The presentation of the reinforcing lights was controlled manually by *E*'s operation of a circuit which was common to all booths. The stimulus input and *S*'s responses were recorded on an Esterline-Angus recorder.

Subjects.—The *S*s were 99 naval aviation trainees in their first week of training.

Procedure.—One of two sets of instructions was given to each *S*. Partial-knowledge (PK) instructions were: "Your job in this test is very easy. You will be taken into another room and each of you will be assigned an individual booth. There are several lights and other equipment in each of these booths. You will only be concerned with two items: a telegraph key and the light located to your left on the panel board in front of you. As I said, your task is quite easy. At the sound of a buzzer you are to guess whether the light will come on or will not come on. If your guess is that it will come on, press the telegraph key. If your guess is that it will not come on, do nothing, just sit there until the next buzzer signal indicating a trial. If you press the key and the light comes on you are correct. If you press the key and the light does not come on, you are incorrect. If you do not press the key you do not find out if you are right or wrong. There will be 10 sec. between each buzzer. Make your choice as soon as possible after the buzzer sounds. Press the key lightly and only one time during a trial." Full-knowledge (FK) instructions and practice procedures were identical except for the elimination of the sentence "If you do not press the key you do not find out if you are right or wrong."

Following a second reading of the instructions, all *S*s were given four practice trials. In the practice trials the light was given in an ABBA order. During the practice trials the recording tape was watched closely for slow or multiple responses. If mistakes were observed, the instructions were reread to the group.

After these instructions and practice trials, the 220 experimental trials were given without a break in timing. Within the two groups, PK and FK, the reinforcing light was keyed to appear on 30%, 50%, or 85%

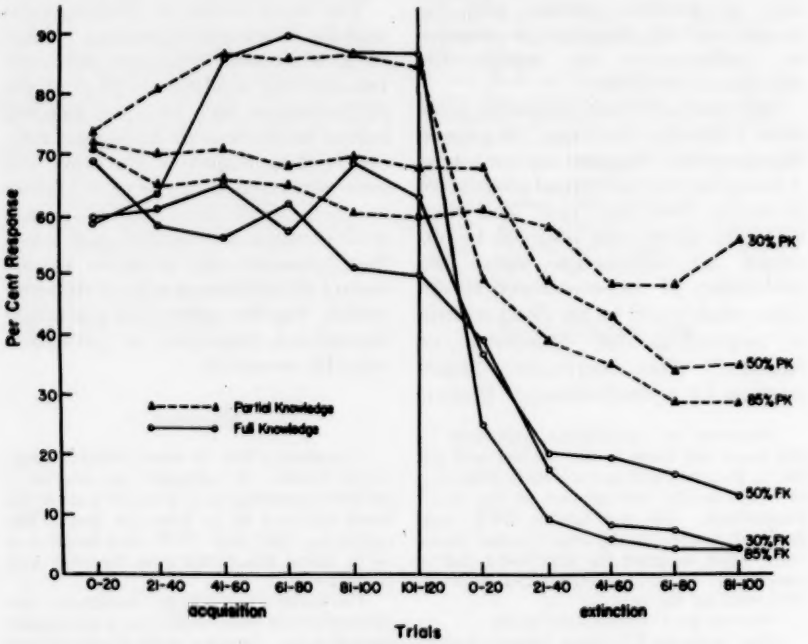

FIG. 1. Acquisition and extinction curves for all experimental groups.

of the trials for the three subgroups, respectively, in a predetermined, essentially random order. The number of *Ss* in each condition was as follows: PK (30%), 17; PK (50%), 18; PK (85%), 15; FK (30%), 11; FK (50%), 27; FK (85%), 11.

On a given trial, a signal was sounded for approximately ½ sec.; this was followed by a period of 9½ sec. before the next signal indicating a trial. In the FK groups a light followed the sounding of the trial warning signal by approximately 4 sec. a predetermined percentage of times. In the PK groups the circuits were set so that if a key was pressed a light would follow the key pressing almost immediately a given percentage of times for the particular group. The same schedule as that used in the FK groups was used in the PK groups except that if the key was not pressed no light would be given on the trial. For all groups, after Trial 120 no reinforcing lights were given.

Results

Figure 1 presents the curves of learning and extinction for the six subgroups.

An analysis of variance was computed using the responses during the last 20 trials of acquisition (Trials 101–120). The results were: for percentage of reinforcement, $F = 25.55$, 2 and 93 *df*; for PK vs. FK, $F = 2.24$, 1 and 93 *df*; and for the interaction of these variables, $F = 1.11$, 2 and 93 *df*. A similar analysis of variance using the responses on Trials 61–80 of extinction was also computed. The *F* ratios, in the same order as the three just cited, were 6.76, 2 and 93 *df*; 67.10, 1 and 93 *df*; and 6.06, 2 and 93 *df*.

We recognize that these analyses should be considered more descriptive than definitive since it was necessary to correct for disparate *N*'s within cells in the manner described by Lindquist (1940), the distributions were skewed, variances were heterogeneous, and the use of the within-groups variance for an error term during extinction may be questioned. Be that as it may, the results reflect those portrayed by the curves: large differences as a function of the ratio of reinforcement during acquisition; very large differences as a function of treatment during extinction.

TABLE 1

PERCENTAGE OF *Ss* (A) RESPONDING ON EXTINCTION TRIALS 81–100 WITH 50% OF FREQUENCY ON TRAINING TRIALS 101–120, AND (B) NOT RESPONDING ON EXTINCTION TRIALS 81–100

Cond.	(A) Responding With 50% Frequency on Extinction Trials 81–100			(B) Not Responding on Extinction Trials 81–100		
	30%	50%	85%	30%	50%	85%
PK	94%	56%	27%	0%	0%	7%
FK	0%	19%	0%	45%	48%	64%

This latter finding is emphasized by the data presented in Table 1, which gives the percentage of *Ss* still responding at 50% as great a rate on Extinction Trials 81–100 as on Training Trials 101–120.

Discussion

Certain data obtained in this experiment may be compared with data obtained by Estes and Straughan (1954). They obtained response rates of 37%, 48%, and 87% during Acquisition Trials 81–120 for reinforcement schedules of 30%, 50%, and 85%, respectively. In contrast, our FK groups, which were similar in the most critical respects to the Estes-Straughan groups, yielded response rates of 50%, 66%, and 87% for the same reinforcement ratios. A considerable elevation of response rates occurred in our experimental situation.

During acquisition the well established direct relationship between ratio of reinforcement and response rate was found (cf. Edwards, 1956; Estes & Straughan, 1954; Grant et al., 1951; Humphreys, 1939; Lewis & Duncan, 1958), but there

were no differences between the PK and FK conditions. However, some factor in our situation, perhaps the press-no press procedure or a difference between our instructions and those of previous studies, resulted in a considerable elevation of the response rates as compared to the findings and predictions of many other investigators, particularly those of Estes and Straughan. This "over response" does not seem attributable to a simple response bias. We ran 18 *Ss* with reverse instructions under the FK conditions. The *Ss* were told to press the key if they did not think the light would appear and not to press the key if they thought the light would appear. A 50% ratio of reinforcement was used. This group's response rate (number of no key presses) in the last 20 trials of training was 71%, a figure not significantly different from that for the 50% FK group in our main experiment.

This elevation of response was anticipated in the restricted-knowledge situation as we felt that the *Ss* would have an "information-seeking" bias. However, the elevation in the FK group was unanticipated.[4]

An inverse relationship between ratio of reinforcement and rate of extinction was clearly obtained in our PK group. This relationship was not obtained in the FK groups, but the group having the highest ratio of reinforcement did extinguish most rapidly. Except for the one reversal in our data there is general support for the discrimination hypothesis as an explanation of the rate of extinction in partial reinforcement designs as related to the training ratios. The hypothesis is treated more thoroughly in the recent paper of Lewis and Duncan (1958).

It would also seem that the initial response rates during extinction would be a factor in determining the rate of extinction. In the present experiment the median rate of response in the 85% groups during the first 20 trials of extinction was 18, in the 50% groups it was 12, and in the 30% groups it was 11.5. Clearly, if the task of *S* during extinction is to learn that a key press is not followed by a light, the 85% group had the greatest opportunity during early extinction.

The primary data of this experiment were the differences in extinction rates noted between the partial-knowledge and full-knowledge training procedures. Part of the difference in responses could be attributed to the different amounts of information furnished *S* on each extinction trial. For the FK group, each trial, regardless of *S*'s response, informed him that guessing "no light" was correct. The no-press response was consistently reinforced for this group. The partial-knowledge *S* obtained information only on those trials on which he responded. It is possible that in the partial-knowledge situation, *S*'s key-pressing response was being reinforced by information at the same time it was being negatively reinforced by not being followed by a light.

The prolongation of extinction in the PK group resembles the extremely persistent behavior obtained in the instrumental conditioning experiments using partial reinforcement with lower organisms. It would appear that such persistence occurs when information

[4] Estes has pointed out the following possibility: "Suppose, however, that *Ss* interpreted NP-NL (no press-no light) as a nonreinforcement, then, applying the methods of the Estes and Straughan paper, we find that the predicted asymptotes should be equal to $1/(2 - \pi)$ where π is the reinforcement ratio. For your 30%, 50%, and 85% groups, the values predicted from this formula are .59, .67, and .87, which you will note are very far from the Estes and Straughan data and quite close to yours. Only the proportion for your FK 30% group is appreciably off; if this is not a sampling deviation, it may indicate that the FK *Ss* are uncertain as to how to interpret NP-NL. If, for example, they interpreted this outcome as signifying that they were "right" and "wrong" with equal probability, then the theoretical asymptotes for the three groups would be approximately .48, .60, and .86." (Personal communication, August 8, 1958.)

can only be obtained by the elicitation of the instrumental response and a failure to respond yields no information.

Summary

For 120 trials *Ss* pressed a key when they guessed a light would come on and did not press the key when they guessed it would not. Three partial-reinforcement ratios were used. In a full-knowledge condition the light appeared or failed to appear on each trial with a fixed probability, independently of *S*'s behavior. In a partial-knowledge condition the *Ss* were told "If you do not press the key you do not find out if you are right or wrong." After the partially reinforced series, 100 additional trials were given with no reinforcement.

No differences were found between informational conditions during training. Significant differences were obtained among reinforcement ratios during both training and extinction. For the partial-knowledge group, extinction was considerably and significantly prolonged.

References

Edwards, W. Reward probability, amount and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–187.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

Lindquist, E. F. *Statistical analysis in educational research.* New York: Houghton Mifflin, 1940.

(Received April 28, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

FRUSTRATIVE FACTORS IN SELECTIVE LEARNING WITH REWARD AND NONREWARD AS DISCRIMINANDA[1]

ABRAM AMSEL AND DAVID L. PROUTY

Tulane University

A study by Amsel and Ward (1954) demonstrated that under certain conditions food reward and nonreward at a choice point can serve as discriminanda for selective learning. Several control experiments were performed in which the more peripheral, mouth-cue aspects of reward and nonreward were reduced by using water instead of food as reward, by introducing a delay at the choice point following drinking or not drinking to allow traces of such exteroceptive cues to dissipate, and by giving *S*s water in the start box a second or two before they found it or did not in the choice-point goal box. These operations did not retard discriminative performance, and the position was taken that frustration drive stimulation arising from nonreward was responsible, at least in part, for the selective learning.

Another consistent feature of the Amsel and Ward results was regarded as supporting a frustration drive stimulus interpretation of their basic finding: Particularly early in the training sequence, the spatial response which was correct following the nonreward discriminandum (the F-side response) occurred significantly more frequently than the opposite (R-side) response. Amsel and Ward offered the interpretation "that when the *S*s begin to respond to the frustration . . . side [the side correct following the nonreward discriminandum], this response is reinforced not only by the reduction of hunger, but also by the reduction of the frustration tension, and a mild fixation of the F-side response develops" (Amsel & Ward, 1954, p. 40).

In the Amsel and Ward experiments, as in the earlier Amsel and Roussel (1952) study, a series of continuously reinforced trials preceded any differential, reward-nonreward testing. In the case of Amsel and Roussel, this testing consisted of measuring and comparing running speed in a runway immediately following a rewarded or a nonrewarded running response; in the case of Amsel and Ward, the test was the selective learning period, requiring *S*s to choose one arm of a T following reward at the choice point and the opposite arm following nonreward. The present experiment varies the reward experience of *S*s in the choice-point goal box preceding the reward-nonreward testing in the Amsel and Ward situation. One group (Group 0) is never rewarded in this goal box prior to the discrimination testing; the other (Group 100) is, like the Amsel and Ward groups, always rewarded in this goal box except on test trials. If the interpretation of Amsel and Ward has merit, there should be an earlier F-side fixation in Group 100 than in Group 0. Studies showing the development of the frustration effect (FE) under partial-reinforcement conditions (Amsel & Hancock, 1957; Roussel, 1952; Wagner, 1959) would make it necessary to predict an eventual FE and subsequent F-side fixation in Group 0.

[1] The preparation of this report was supported by a grant, G-5527, from the National Science Foundation.

If frustrative factors are involved in the reward-nonreward selective learning, it should be possible to show, following Amsel and Roussel (1952) and those mentioned immediately above, that running from the goal box at the choice point (G_1) to either of the two goal boxes at the ends of the T arms (G_L, G_R) is more vigorous following nonreward than following reward in the choice-point goal box. On the basis of the frustration hypothesis one might expect a more immediate indication of the FE in Group 100 than in Group 0.

METHOD

Subjects and Apparatus

The *S*s were 24 naive, male, hooded rats from the colony at Newcomb College, Tulane University. The approximate age of the *S*s was 5 mo. at the start of preliminary training.

The apparatus, except for the modification necessary for timing the running response, has been described elsewhere (Amsel & Ward, 1954). It consisted of two parts which shall be designated A and B. Part A was made up of a gray start box, 12 × 4 in.; a straight, gray runway (Runway 1), 42 × 4 in.; a white goal box (G_1), 12 × 4 in.; a second gray runway (Runway 2), 15 × 4 in.; and a second black goal box (G_2), 12 × 4 in. All of these were connected in series. Apparatus B was a single unit T maze. It consisted of a stem portion (start box, runway, and G_1 of Part A) and a cross portion to complete the T. This cross portion consisted of a gray alley, 15 × 4 in., with black goal boxes at both ends.

A photoelectric timing system operating with infrared light was used to measure running time in Apparatus B. Interruption of an infrared beam just outside of the guillotine door of G_1 started a .01-sec. Standard Electric timer; the timer was stopped when *S* broke the infrared light beam just outside either terminal goal box in the arms of the T. A .1-sec. stop watch was used for other timing.

Procedure

The experimental period covered 64 days.

Preliminary training.—This was a period of 10 days during which *S*s were handled by *E* and adjusted to a 20-hr. water deprivation schedule according to a procedure previously described (Amsel & Ward, 1954).

Adjustment to the apparatus.—This was a 2-day period. On the first day, *S* was placed in the start box with the door closed; all other doors were left open. The *E* then raised the door of the start box and allowed *S* to explore the entire T for 10 min., during which time the photoelectric timing system was on to adjust *S* to the mild clicking sounds of the relays housed in sound-deadened boxes. The same procedure was repeated on the second day using the straight runway (Apparatus A). Preliminary training to the swinging doors and to Apparatus A followed the Amsel and Ward (1954) procedure.

Introduction of the variable.—At this stage, *S*s were divided into two groups of 12 *S*s each using the method of litter-mate controls. Apparatus A was used during this part of the procedure. Group 100 always received two drops of water in G_1 and in G_2. Group 0 always received two drops of water in G_2 but never received water in G_1. All *S*s were given 10 trials a day for 5 days (50 trials) under this procedure.

Selective learning.—This procedure was followed for 43 days. On each day, *S*s were first given two trials in Apparatus A following the previously outlined procedure. The *S*s of Group 100 were always rewarded on G_1 on these trials, *S*s of Group 0, never. Both groups were always rewarded in G_2. This was followed by four trials in Apparatus B. The procedure was then repeated so that eight trials of selective learning and four trials of straightaway training were run each day. On the Apparatus B trials, *S* was either rewarded in G_1 or not rewarded in G_1 according to a prearranged order. For any particular *S*, for example, choice-point reward followed by response to the right led to further reward, while the same spatial response following choice-point nonreward led to further nonreward. Choice-point reward followed by response to the left, of course, led to nonreward, etc. Noncorrection was employed throughout.

When *S* found water in G_1, it was released 2–5 sec. after drinking was completed to the choice point of the T. The time taken to drink the water was of the order of 5 sec. On the trials when *S* found no water in G_1 the interval in G_1 was varied randomly between 7 and 10 sec. to ensure that a temporal differentiation would not be established between reward and nonreward trials. The delay on reward trials was to allow for dissipation of peripheral cues resulting from water

in mouth. The order of nonrewards (F) and rewards (R) used in G1 was RFFR FRRF and FRRF RFFR on successive days for half the *S*s. This order was reversed for the other half of each group. The *S*s were run in rotation, and approximately 15 min. separated successive trials for the same *S*.

The direction of the response after leaving the choice point in Apparatus B was recorded, and running time was measured from just outside G_1 to just outside whichever goal box *S* entered.

Results

The F-side fixation.—Figure 1 shows the mean number of correct responses for Group 0 and Group 100, as well as the mean number of responses to the F-side in each group. The F-side fixation becomes noticeable in Group 100 at about Day 11. The F-side curve for this group rises sharply and maintains a level of about 70% for about five days (40 trials). It then declines as the number of correct responses increases. The curve showing number of responses to the F-side for Group 0 shows a sharp drop (an R-side fixation) which corresponds to the rise shown by the Group 100 curve. The curve showing correct performance for Group 0 is consistently above the corresponding curve for Group 100 with the exception of two reversals. Mean errors made by Groups 0 and 100 were 92.3 and 118.8, respectively. This difference was significant beyond the .05 level ($t = 2.53$).

For each group a t test of the difference in number of R-side and F-side responses was performed. A t test was also run of the difference between the R-F differences in the two groups. For this analysis the data were arbitrarily divided into two 21-day segments (Day 1 was excluded). The first segment (Days 2–22) in-

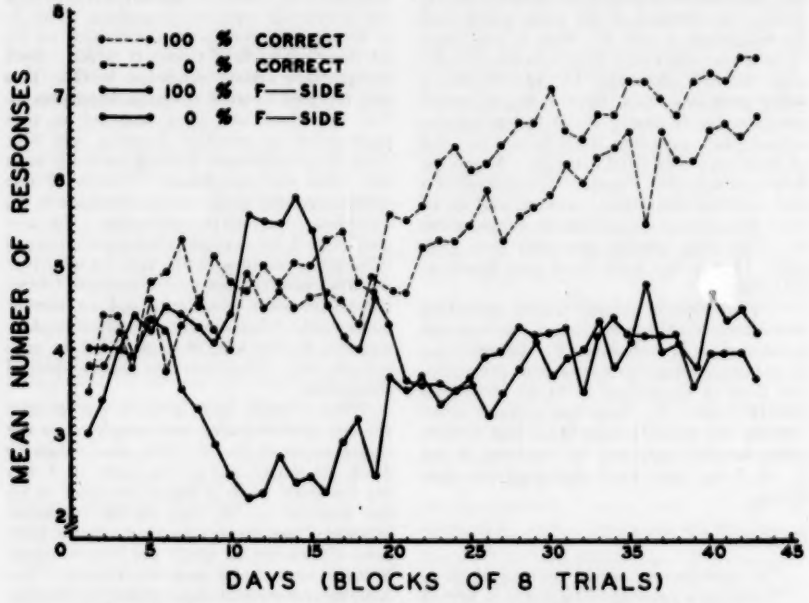

FIG. 1. Correct responses and F-side responses for Groups 0 and 100.

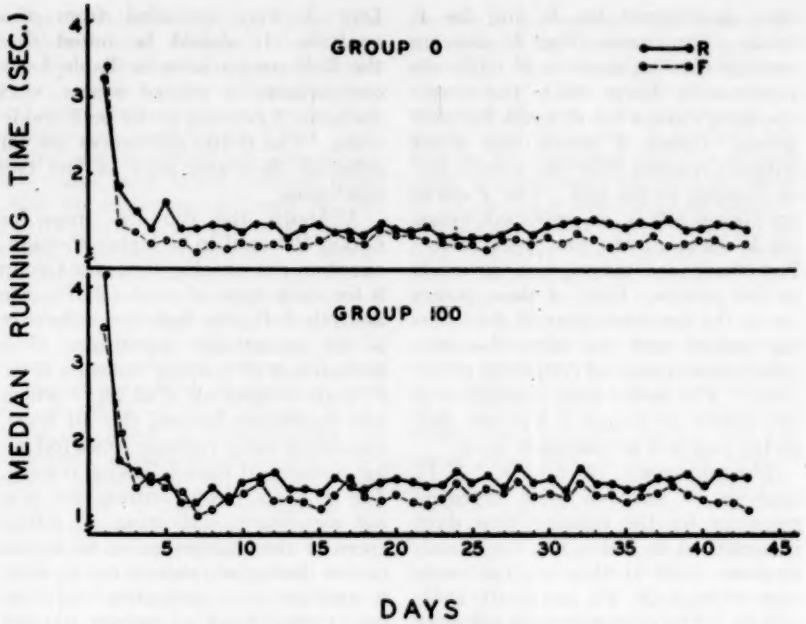

FIG. 2. Running time following F and R in G_1 for Groups 0 and 100.

cludes the period of the F-side fixation by Group 100 and the R-side fixation by Group 0. The second segment (Days 23–43) begins approximately at the stage where the number of correct responses increases sharply for both groups. The results indicate a number of runs to the R-side significantly greater than chance ($t = 2.39$, $P < .05$) for Group 0 during the first half of the training period. For Group 100 the difference was in the opposite direction (i.e., a greater number of runs to the F-side) although not significant ($t = 1.51$). It is felt that this failure to show significance for Group 100 by this test was due largely to one *S* in the group which developed a strong R-side fixation which may have been a position preference. Also, for both the Group 0 and Group 100 analyses, the tests were conservative in that the Day 2–22 segment included the days before the fixations developed and also those on which they started to decline to the chance level as the number of correct responses increased. During the last 21 days there was no significant difference in the number of R-side vs. F-side runs for either group.

A comparison for Days 2–22 of the R-side vs. F-side frequency difference scores for Group 0 with those for Group 100 showed that these differences, which are in opposite directions are significantly different at about the .01 level ($t = 2.76$).

The frustration effect.—Daily median running times from the choice point to either G_L or G_R were obtained for each *S* on R and on F trials. An over-all median of such individual-*S* medians on each day was

then determined for R and for F trials. The curves (Fig. 2) showing median running time on F trials are consistently lower than the corresponding curves for R trials for each group. Group 0 shows this effect without reversal from the second day of training to the end. The *F* curve for Group 100 is, as predicted, below its R curve at the first point (day) but there are subsequent reversals at five points. Four of these points are in the first nine days of the training period and the effect becomes more pronounced and consistent thereafter. The earlier final separation of the curves in Group 0 was not predicted and will be discussed later.

The summary of a $2 \times 2 \times 12$ analysis of variance with repeated measures for the running time data is presented in Table 1. The classifications used in this analysis were type of trial (R, F), group (0, 100), and *Ss*. The entries in each cell were seven median running times for the seven successive six-day blocks of trials. These were medians of either six daily F medians or six daily R medians for each *S*. The data of Day 1 were excluded from this analysis. It should be noted that the R-F comparisons in Table 1 are comparisons of related scores, with the same *S* running under both conditions. The 0–100 differences are for different *Ss* under each of the two conditions.

TABLE 1

ANALYSIS OF VARIANCE OF INDIVIDUAL *S* MEDIAN RUNNING TIMES

Source	*df*	*MS*	*F*
Groups (0, 100)	1	.963	.20
Discriminanda (R, F)	1	1.979	28.27**
Between *Ss*	22	.803	
Blocks of days	6	.212	3.03**
Blocks × 0, 100	6	.062	
Blocks × R, F	6	.037	
0, 100 × R, F	1	.001	
Blocks × 0, 100 × R, F	6	.093	1.33
Pooled *Ss* × R, F	22	.053	
Pooled *Ss* × Blocks plus Pooled *Ss* × Blocks × R, F	264	.070	
Total	335		

** $P = .01$.

Although the running times for Group 100 tend to be higher throughout than the running times for Group 0 for each type of trial (R, F), the analysis indicates that this difference is not statistically significant. The comparison of running times on R vs. F trials yielded an *F* of 28.27 which was significant beyond the .01 level, indicating faster running times following nonreward than following reward. The Days × Group interaction was not significant, indicating no differences in the changes going on in the curves during acquisition or, to state it another way, indicating that the four curves tend to remain parallel throughout discrimination training. The between-days variance was significant beyond the .01 level ($F = 3.03$). This simply indicates a decrease in running times for all groups over days of training, reflecting mainly the difference between level of performance on the first few days and all the rest.

DISCUSSION

The F-side fixation.—The fixation of the F-side response by Group 100 confirms the finding of Amsel and Ward with a comparable condition and, by itself, it could be interpreted, with them, as representing a fixation effect due to the reduction of both thirst and frustration on the F-side as against thirst reduction alone on the opposite side of the T. However, the appearance of the opposite (R-side) fixation in Group 0, instead of the later F-side fixation that was predicted, does not support the Amsel and Ward interpretation of the F-side phenomenon.

The authors recognized, when the experiment was designed, that there is an interpretation for the nonpredicted finding which stresses stimulus change and associative mechanisms rather than motivational mechanisms. This associative interpretation can account for the Amsel and Ward fixation effect under the conditions of Group 100 and also for the Group 0 finding of the present study. The important stimulus change for Group 100 is the introduction of nonreward following a preliminary period of consistent reward in G_1. For Group 0 the important stimulus change in G_1 is the introduction of reward following a preliminary period of consistent nonreward. Both of these changes might have the effect of compounding the differential stimulation in G_1 and at the choice point after the change. This should lead to faster learning of the response which is correct following the compound stimulus (Hull, 1943, Ch. 13). It is possible also that both drive-incentive factors and associative factors may be operating in both groups to effect the fixations which are apparent. In this case, the associative factors would seem to play a more important role in Group O, producing an R-side rather than an F-side fixation. The role of these two kinds of factors in discriminations based upon rewards and nonrewards as discriminanda would seem important enough to warrant further experimentation to attempt to separate and isolate their effects.

The frustration effect.—In both Group 0 and Group 100 the FE was observed, i.e., running was faster on F trials than on R trials of the discrimination. What was not predicted, however, was the appearance of this difference on the first trials of selective learning in Group 0 and its failure to appear from the very outset in Group 100. This failure of the FE to appear on the earliest test trials is not new in our studies, although it does not always happen. In the earliest study in this series (Amsel & Roussel, 1952) the FE took a few trials to develop. An explanation was briefly suggested then which is interesting, if not persuasive, in this context. This is basically the same as the associative explanation offered for the F-side and R-side fixations although its effect is different. Again stimulus change, from nonreward to reward and vice versa, is involved but now we are concerned not with the cue value of the stimulus but with the effect of the change of stimulus on an already-established, immediately-following running response. This ordinarily comes under the heading of stimulus generalization decrement or, in Pavlovian terms, external inhibiting effects.

An interpretation in terms of generalization decrement accounts for both the early difference in Group 0 and the lack of it in Group 100. On the early selective learning trials, Group 0, which has never found food in G_1, now finds it there on half of the trials; Group 100, which has always found food in G_1, now finds no food there on half of the trials. We might expect responses following reward in G_1 to be inhibited (slower than they should be) for Group 0, while responses following nonreward in G_1 should be inhibited in Group 100. We are suggesting that there are two factors —one associative and one motivational— contributing to the separation of the R and F curves for Groups 0 and 100; that one of these factors, generalization decrement, produces a large pseudo-frustration effect in the early selective learning trials of Group 0 because it increases running time following reward in G_1, and produces a pseudo-nonfrustration effect in the early trials of Group 100 because it increases running time following nonreward in G_1. Examination of Fig. 2 will show why this explanation suggests itself: (*a*) the separation of the Group 0 curves, after the first day, is large at the beginning, then gets successively smaller over the first 20 days of selective learning, then larger again; and (*b*) there are five points at which the F running time curve is higher than the corresponding R curve in Group 100 and these are all within the first 20 points, the largest inversion being at the second point of the curve.

Summary

This investigation was a follow-up of an experiment performed by Amsel and Ward (1954) which dealt with the implications of ascribing certain motivational properties to frustration in a selective learning situation. The present experiment introduces an independent and a dependent variable not involved in the Amsel and Ward study: the manipulation of reward experience prior to the discrimination training, and the measurement of running time after reward and after nonreward in the discrimination trials. Using 24 hooded, male rats as *Ss* in a T maze similar to that employed by Amsel and Ward, with reward and nonreward as the discriminanda, eight discrimination trials were run each day for 43 days. Thirst was employed as the relevant drive.

The results indicate that in Group 100 (previous history of consistent reward at the choice point), which is comparable to the basic Amsel and Ward condition, the response which is correct following nonreward at the choice point (F-side response) tends to be fixated early in training relative to the response which is correct following reward (R-side response). In Group 0 (previous history of nonreward at the choice point) an R-side fixation was observed.

For both Group 0 and Group 100 there was greater vigor of response following nonreward at the choice point than following reward. This effect shows up earlier in Group 0 than in Group 100. Implications of the two basic findings are discussed.

References

Amsel, A., & Hancock, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

Amsel, A., & Roussel, J. S. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

Amsel, A., & Ward, J. S. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–47.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Roussel, J. S. Frustration effect as a function of repeated nonreinforcements and as a function of the consistency of reinforcement prior to the introduction of nonreinforcement. Unpublished master's thesis, Tulane Univer., 1952.

Wagner, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

(Received May 15, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

SWEETNESS AND SALTINESS OF COMPOUND SOLUTIONS OF SUCROSE AND NaCl AS A FUNCTION OF CONCENTRATION OF SOLUTES

J. G. BEEBE-CENTER, M. S. ROGERS,[1] W. H. ATKINSON,[2] AND D. N. O'CONNELL

Harvard University

Available evidence indicates that the sweetness of a compound aqueous solution of sucrose and NaCl is greater than that of the constituent sucrose solution presented alone when the concentration of NaCl does not exceed 0.21% (Fabian & Blum, 1943; Kremer, 1918), less for higher concentrations of NaCl (Hambloch & Püschel, 1928; Heymans, 1899), and that its saltiness is less than that of the constituent NaCl solution presented alone when the concentration of sucrose is 1% (Fabian & Blum, 1943) or higher (Hambloch & Püschel, 1928; Heymans, 1899). The data are too limited, however, to allow any but the vaguest inferences concerning the general functions relating sweetness and saltiness of the compound solution to concentration of the two solutes. The experiment described below was undertaken to provide the data necessary for a first approximation of these general functions.

METHOD

The compound solutions used in the experiment were the 25 possible binary combinations of −.52, −.02, .48, .98, and 1.48 log gm. NaCl and .00, .50, 1.00, 1.50, and 2.00 log gm. sucrose per 100 cc. tap water. The term combination here refers to solutes, not solutions. Thus −.52 log gm. NaCl and .00 log gm. sucrose poured into a graduate containing 100 cc. of tap water represents a compound solution of −.52 log gm. NaCl and .00 log gm. sucrose per 100 cc. tap water. Choice of logarithmic steps was based on the tendency of $\Delta I/I$ to be constant in taste (Pfaffmann, 1951, p. 1156). Choice of concentrations represented an attempt to cover as fully as possible the ranges from threshold to solubility limit at 70° F for NaCl, namely −.70 (Pfaffmann, 1951, p. 1152) to 1.577 (Hodgman, 1933, p. 751) log gm. NaCl per 100 cc. tap water, and for sucrose, namely −.16 (Pfaffmann, 1951, p. 1152) to 2.505 (Hodgman, 1933, p. 771) log gm. sucrose per 100 cc. tap water.

Two *Ss* were used in the experiment. In a first part, the sweetness of the compound solutions was measured in terms of the log concentration (log gm. sucrose per 100 cc. tap water) of the equi-sweet simple sucrose solution. The equi-sweet sucrose solution was determined by one of two methods, both of which involved use of a scale of 16 sweetness standards, starting with pure tap water,[3] and then extending from −1.25 to 2.25 log gm. sucrose per 100 cc. tap water in .25-log steps.

For the nine compound solutions representing the nine binary combinations of −.02, .48, and .98 log gm. NaCl and .50, 1.00, and 1.50 log gm. sucrose per 100 cc. tap water, the method used to determine the equi-sweet simple sucrose solution was what Dixon and Massey (1951, pp. 278–286) call the "up and down" method. Instead of using the mean as the measure of central tendency we used the median. Furthermore, when the compound was judged less sweet than the least concentrated standard of the scale, or sweeter than the most concentrated, the same standard was used over again in the next comparison. Medians were based on 36 judgments for each compound distributed over nine 1-hr. test periods. Each test period involved four judgments on each of the nine compounds presented in chance nonrecurrent series. In a first group of nine trials the compound was always presented

[1] Now at System Development Corporation, Santa Monica, California.

[2] Now at U. S. Naval Medical Research Laboratory, New London, Connecticut.

[3] This standard was treated in computations as though it had been one of −1.25 log gm. sucrose per 100 cc. tap water.

first, the simple sucrose standard second. The order was reversed in a second group of nine trials. The solutions were presented in 1-oz. glasses. The *S*s were instructed to sip about 10 cc. of solution from a first glass, expectorate, rinse, wait 20 sec., sip from a second glass, expectorate, rinse, and state whether the second was more or less sweet than the first.

For the other 16 compound solutions and for 5 simple control solutions of .00, .50, 1.00, 1.50, and 2.00 log gm. sucrose per 100 cc. tap water, the method used was a simple matching method. The *S* had before him a series of 16 glasses constituting the sweetness scale mentioned above. The standards were in order of concentration and labeled with respect to log concentration. His instructions were simply to choose by taste the standard most nearly approximating the compound with respect to sweetness. He was completely free with respect to tasting procedure. Matching usually involved 5 to 10 samplings of the compound (or control) and of various standards. Each of three sets of seven compounds (or controls) was presented for matching in two nonrecurrent chance series on two successive days. Thus the six daily sessions devoted to matching yielded four matches for each of the 16 compounds (or controls).

Saltiness of the compound solutions and of five simple control solutions of −.52, −.02, .48, .98, and 1.48 log gm. NaCl per 100 cc. tap water was measured in a second part of the experiment in terms of the log concentration (log gm. NaCl per 100 cc. tap water) of the equi-salt simple NaCl solutions. The method used was in all cases the direct matching method described above. The set of salt standards started with pure tap water and then extended from −2.02 to 1.48 log gm. NaCl per 100 cc. tap water in .25-log steps.

Results

Figure 1 presents the results of the sweetness measurements averaged (in log terms) for the two *S*s. Each circle represents a particular compound solution or control solution. Circles representing solutions with the same sucrose concentration are connected by a line, with the concentration of sucrose marked (in log gm. solute per 100 cc. tap water) above the line. Abscissae represent NaCl concentration of the compound (or control) solutions. Ordinates represent sweetness in terms of the concentration of an equi-sweet simple sucrose solution. Figure 2 presents the results of the saltiness measurements in like manner.

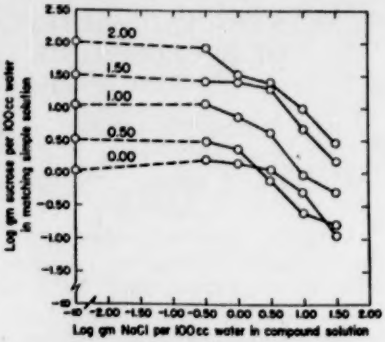

FIG. 1. Sweetness of compound solutions of sucrose and NaCl (ordinates), for five different levels of sucrose concentration (lines), as a function of five different concentrations of NaCl (abscissae).

The major effect shown in Fig. 1 and 2 is clearly one of masking. Figure 1 does indeed show enhance-

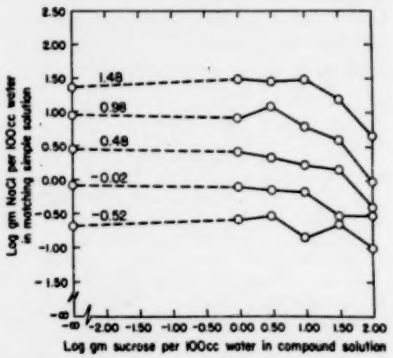

FIG. 2. Saltiness of compound solutions of sucrose and NaCl (ordinates), for five different levels of NaCl concentration (lines), as a function of five different concentrations of sucrose (abscissae).

ment of the sweetness of a 1% (.00 log %) concentration of sucrose by addition of up to 3% (.48 log %) of NaCl, and also of a 10% (1.00 log %) concentration of sucrose by addition of NaCl to yield a .3% (−.52 log %) concentration. The amount of enhancement, however, is slight when compared with the amount of masking by high concentrations of NaCl.

Beebe-Center, Rogers, and Atkinson (1955) have shown that with respect to subjective intensity the effectiveness of sucrose solutions (in percentage concentration) is about 1/10 that of NaCl solutions (in percentage concentration). We can thus specify a sucrose solution in terms of the intensively equivalent NaCl solution by multiplying the concentration of the sucrose solution by .1. If, now, we specify the masking solutions used in the present experiment, both sucrose and NaCl, in terms of intensively equivalent NaCl solutions, the averages of our data for all five parameter values can, in the case of *both* Fig. 1 and 2, be fitted reasonably well by a *single* mathematical expression. The expression is $\log (C/C_o) = -0.159M^{.7}$, in which C_o and C represent respectively the masked solution in the compound and the like simple solution that matches it, both in grams of solute per 100 cc. tap water, and M represents the masking solution in the compound in grams of solute per 100 cc. tap water of the intensively equivalent NaCl solution. The degree of fit is shown in Fig. 3.

An equally good fit can be obtained if M is specified in terms of the osmotically equivalent NaCl solution (the factor multiplying sucrose concentration is in this case .094). Indeed, it was with this specification that we originally developed the equation above. Specification based on osmotic equivalence, however, seems to promise less generality for our equation than does specification based on intensive equivalence. Preliminary observations indicate that suprathreshold concentrations of quinine sulfate have appreciable masking effects, although their osmotic effectiveness is very slight. Furthermore, a laboratory exercise conducted on each other by G. Bermant and S. Speeth in the Harvard Laboratory showed that the masking effect of a glucose solution on the salinity of an NaCl solution is less than that of the osmotically equivalent but sweeter sucrose solution.

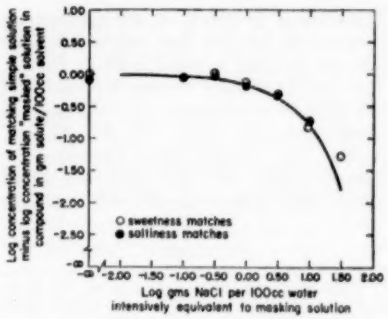

FIG. 3. Change in taste of masked component of compound solutions of sucrose and NaCl (ordinates) as a function of concentration of NaCl solution intensively equivalent to masking component (abscissae). Change in taste is specified in terms of the logarithm of the ratio of the concentration of the matching simple solution to the concentration of the like masked solution. Circles represent averages of changes in sweetness for all five levels of sucrose concentration of the compound solutions. Dots represent averages of changes in saltiness for all five levels of NaCl concentration of the compound solutions. The line represents the expression $\log C/C_o = -.159M^{.7}$ in which $\log (C/C_o)$ stands for the ordinate of the figure and M for the abscissa.

The exercise by Bermant and Speeth, it should be added, was done in part to check on the representativeness of one of the points in Fig. 2. The point in question was the second from the right on the line labeled .48. The mean ordinate value for the two *S*s was .17. The combined ordinate value for Bermant and Speeth was .08. Exercises were also conducted with P. Derks, L. Gollub, H. Rubin, B. Scharf, M. Taylor, R. Teghtsoonian, J. Urban, and S.

Wasserman to check on the representativeness of the first five points from the left on the line labeled 1.00 in Fig. 1. The mean ordinate values in order from left to right for the two *S*s in the present experiment were 1.03, 1.06, .87, .63, and −.02. The corresponding combined ordinate values for the eight *S*s listed above were 1.04, 1.01, .90, .58, and .08.

Summary

The sweetness and saltiness of 25 compound solutions of sucrose and NaCl were measured in terms of the concentration of equi-sweet simple solutions of sucrose and of equi-salt simple solutions of NaCl. The compound solutions sampled in steps of one-half log unit cover almost the entire ranges of concentrations from threshold to solubility limit for sucrose and NaCl.

Some enhancement of sweetness was found in the case of weak solutions. The principal effect, however, was one of mutual masking. Amount of masking, averaged over different concentrations of the masked solution, is well described by a single expression in the case both of masking of sweetness by NaCl and of masking of saltiness by sucrose. The expression is $\log (C/C_0) = -.159 M^{.7}$ in which C_0 and C represent respectively the masked solution in the compound and the like simple solution that matches it, both in grams of solute per 100 cc. tap water, and M represents the masking solution in the compound, in grams of solute per 100 cc. tap water of the intensively equivalent NaCl solution.

References

Beebe-Center, J. G., Rogers, M. S., & Atkinson, W. H. Intensive equivalences for sucrose and NaCl solutions. *J. Psychol.*, 1955, **39**, 371–372.

Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.

Fabian, F. W., & Blum, H. B. Relative taste potency of some basic food constituents and their competitive and compensatory action. *Food Research*, 1943, **8**, 179–193.

Hambloch, H., & Püschel, J. Ueber die sinnlichen Erfolge der Darbietung von Geschmackemischungen. *Z. Sinnesphysiol.*, 1928, **59**, 136–150.

Heymans, G. Untersuchungen über psychische Hemmung. *Z. f. Psychol. u. Physiol. d. Sinnesorgane*, 1899, **21**, 321–359.

Hodgman, C. D. (Ed.) *Handbook of chemistry and physics.* (18th ed.) Cleveland: Chemical Rubber Publ. Co., 1933.

Kremer, J. H. Influence de sensations du goût sur d'autres spécifiquement différentes. *Arch. néerl. de physiol. de l'homme et des animaux*, 1918, **1**, 625–634.

Pfaffmann, C. Taste and smell. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1143–1171.

(Received May 5, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

THE GELB EFFECT [1]

EDWARD C. STEWART [2]

University of Texas

A phenomenon which has been important in raising problems concerning achromatic colors was reported in 1929 by Gelb. He conducted his demonstration in a half-darkened room lighted by an overhead lamp. Hidden from *O* was an arc lamp so arranged that its cone of light was coincident with the boundaries of a spinning disk. The spinning disk was velvet-black under ordinary conditions of viewing, but was perceived as "white or whitish-gray." When a strip of white paper was held in front of the disk it turned black. The white paper was removed and the disk again became white. When the rotating disk was stopped, it was perceived as black. The *O* perceives the spinning disk as white even with full knowledge of the nature of the demonstration.

Various authors have discussed the Gelb effect, and some have attempted an explanation of the phenomenon. Gelb originally cited his experiment as evidence that at least two surfaces of different albedos must be present in the field for a perception of illumination and illuminated object (Gelb, 1929). After describing his experiment, Gelb compares his effect with the effect obtained by Hering in the latter's shadow experiment. In each case, an experimental variable changes the organization of the field of vision by introducing a new surface with a different intensity. In Gelb's experiment the experimental variable was the strip of white paper, and in Hering's experiment, a contour that covered the penumbra of a shadow. Gelb ends his discussion by pointing out that each achromatic surface has a constant reflectance and hence maintains its rank order among other reflectances as illumination is changed. Although Gelb is not explicit on the point, the above statements indicate that Gelb's explanation of the phenomenon would be a gradient theory. The achromatic color perceived is relatively independent of the illumination, and hence the perception of a surface would seem to depend on the relation of the albedo of the observed surface to all other surfaces in the visual field. The relation of the albedo of the observed surface to all other intensities in the visual field is undoubtedly what Gelb means by organization of the visual field.

Kardos (1934), following Gelb, stresses the differences in the albedos of the surfaces observed, and emphasizes the structure of the field of vision. He offers no specific explanation of the phenomenon, but indicates a need for more experiments and theory. For Katz (1935), the explanation of the Gelb effect pivots on the difference between film and surface colors. The strip of paper precipitates the surface mode of viewing by introducing degrees of distinctness of material structure and the insistence of the field. The perception of microstructure is also a reason why the disk is seen as black when not rotating.

Koffka's (1935, pp. 245–247) theory

[1] This paper is based on a doctoral dissertation presented to the University of Texas. The writer gratefully acknowledges the supervision and guidance of H. Helson.

[2] Now at Lehigh University.

is quite different from that of Katz. He emphasizes the presence of a difference between the intensities furnished by the room itself, and by the disk during the conditions of the original demonstration. If the ratio of the two intensities is in the neighborhood of 60:1, the room will be seen as black and the disk as white. This prediction is in conformity with the facts of the experiment and of the lightness ratio between the room and the disk. When the strip of paper is introduced in front of the disk, the stimulus gradations of lightness are determined by three intensities: the room, the disk, and the paper. The disk may now remain white, since the same relation between it and the black room still holds, or it may turn black as a function of the relation between the disk and the paper. In reality, the disk turns black because of "appurtenance" . . . "the white strip of paper belongs more closely to . . . the disk than to the background . . ." (Koffka, 1935, pp. 246–247).

Osgood (1953) as well as Woodworth and Schlosberg (1954) take exception to any explanation of the Gelb phenomenon that is even partially dependent on contrast, and hence they object to the theory of Koffka. These writers maintain that the area of the paper is not sufficient to induce contrast and point out that the paper probably serves as a cue to illumination.

The purpose of the present study is to examine the parameters that affect the Gelb phenomenon. Since this investigation was conceived within the framework of Helson's (1947) adaptation-level theory, contrast was chosen as the principle most likely to yield a satisfactory understanding of the Gelb effect (Helson, 1938, 1943). If contrast is the explanation, then this phenomenon is not an all-or-none effect, and it should be subject to variations and to quantitative determination. The principle of contrast also suggests that two important parameters are the size and position of stimuli seen with the Gelb disk (Diamond, 1955; Wallach, 1948). Specifically, the present investigation is concerned with the verification of the following hypotheses:

1. The Gelb effect is a contrast phenomenon and subject to the usual laws of lightness-contrast.

2. The lightness of the Gelb disk is a function of all the reflectances in the field of view, and will be modified by change of reflectance of any part of the field (Henneman, 1935; Judd, 1941). The greatest modifications of the Gelb disk will occur with changes in reflectance of objects in the area of the disk itself.

3. The effect of any reflectance will vary with the size of the reflecting object and with its distance from the disk. The closer the object, and the larger its size, the greater will be the effect of its reflectance on the disk.

In accordance with Hypothesis 2, the introduction of any object, no matter how small, should have an effect on the lightness of the disk, in line with Judd's formula for lightness (whiteness, see Judd, 1941), which makes whiteness dependent on the ratio between the highest and lowest reflectances in the field.

Method

Apparatus.[1]—Each *S* was presented the stimuli in the form of two disks which were rotated at a speed sufficient to eliminate characteristics of their surfaces. On *S*'s right

[1] The writer gratefully acknowledges the assistance of Karl N. Vartia in designing and building the apparatus.

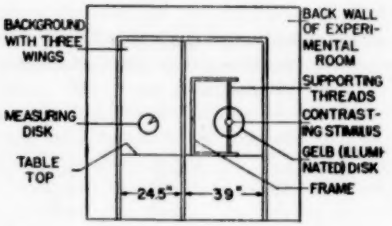

FIG. 1. The *S*'s view of the apparatus.

was the black, brightly illuminated disk, and on *S*'s left was the standard measuring disk (Fig. 1 and 2). The brightly illuminated disk, constructed of masonite, was 12 in. in diameter and painted a flat black. The measuring disk consisted of two interlocking disks made of light-weight aluminum, each having a diameter of 8 in. One of these disks was painted black to match the Gelb disk and the other was painted white. The two aluminum disks were used conjointly to produce variations in degrees of black and white from 0° to 360°. The contact-nut in the center of the measuring disk was covered by black tape.

The two disks were displayed against a background of black cloth. The cloth-covered frame measured 63.5 in. in width and 95 in. in height, and had three wings attached at right angles to the background frame. The wings extended 57 in. toward *S*, with one wing at each end of the background and the third wing separating the two disks. The partition thus created for the Gelb disk was covered on top by black cloth, while the top of the partition for the measuring disk was left uncovered. The table on which the supports for the disks rested was draped with black cloth; hence, both disks appeared against a dark background.

The contrasting stimuli were cut from construction paper and covered with a white paint which matched the white of the measuring disk. Three sizes of contrasting stimuli were used: circles of ½, 1, and 2 in. in diameter. A black wooden frame in the shape of a squared "C" was constructed for each size of contrasting stimulus and between the open ends of this frame the stimuli were suspended by thin black thread. The frame slid along a board which was tightly clamped to the table top. Neither the frame nor the threads were visible to *S*; however, the threads cast small and distinct shadows on the surface of the disk. Since *S*'s line of view and the trajectory of the light were nearly the same, the shadows of the contrasting stimuli were hardly noticeable. Holes were drilled in this board and in the frame so that by aligning the proper holes in the board with those in the frame with a brass peg, the contrasting stimulus could be placed at any desired position in front of the Gelb disk.

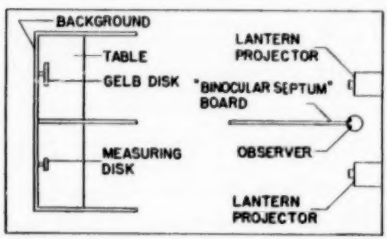

FIG. 2. Schematic layout of the apparatus.

Illumination.—Each disk was illuminated by a separate lantern projector. The right, or Gelb disk, received the cone of light coincident with its boundaries. The measuring disk, however, lay in a rectangular field of lighted surface, the bottom boundary of which coincided with the table top and included 65.5 in. of the background extending upward. Some of the light also fell forward on the two wings to a distance of 21 in.

The illumination of the measuring disk was determined for each *S* in a series of trials preceding the main experiment. Each *S* was shown the two conditions of the Gelb disk in isolation and the Gelb disk with the largest contrasting stimulus in the center. A level of illumination was chosen that allowed for an appreciable difference between the two extreme conditions of the Gelb disk, and yet enabled each condition to be matched with the measuring disk. It was eventually feasible to conduct the experiment with the same illumination for three of the four *S*s (Table 1).

Procedure.—The *S* was seated approximately 15 ft. in front of the disks, with his face pressed against the edge of a black up-

TABLE 1

SPECIFICATION OF LUMINANCE VALUES

*S*s	Gelb disk	Measuring Disk	
		360° Black	360° White
DV, JMc, JM	.45 ft.-L.	.022 ft.-L.	.61 ft.-L.
JF	.45 ft.-L.	.011 ft.-L.	.30 ft.-L.

TABLE 2

POSITION OF CONTRASTING STIMULI IN RELATION TO THE GELB DISK

Stimuli	Diameter	C	Position[a]			
			1	2	3	O-T
Small	0.5	0	1.938	3.813	5.689	6.25
Medium	1.0	0	1.813	3.688	5.500	6.50
Large	2.0	0	1.656	3.344	4.938	7.00

Note:—Both halves of the Gelb disk were similar in reference to the contrasting stimuli.
[a] All measurements of position are given in inches from the center of the Gelb disk to the center of the contrasting stimulus.

right board, which acted as a binocular septum. Before each series of trials, the board was adjusted to allow only the left eye to see the measuring disk, and only the right eye to see the Gelb disk.

The Gelb disk was presented first, and then followed by the measuring disk. The presentation of each disk lasted for 3 sec. The *S* closed his eyes after each trial while waiting for the next trial. The method of limits was used throughout the experiment.

All four *S*s were tested with the same experimental design. Each contrasting stimulus was first presented in the center and then moved to the right through three successive positions to the tangent position on the outside of the Gelb disk (center to right-outside positions; see Table 2). The subsequent series were right-outside-tangent (R-O-T) to center (C), left-outside-tangent (L-O-T) to C, and C to L-O, the design thus furnishing two measurements for each of the nine positions used. If one considers the right and left sides of the Gelb disk to be equivalent, then each of the five positions yields four measurements for each *S* and for each size of the contrasting stimulus.

RESULTS

The results of the experiment are reported with all readings of the measuring disk converted to degrees of black. Figure 3 has been plotted to

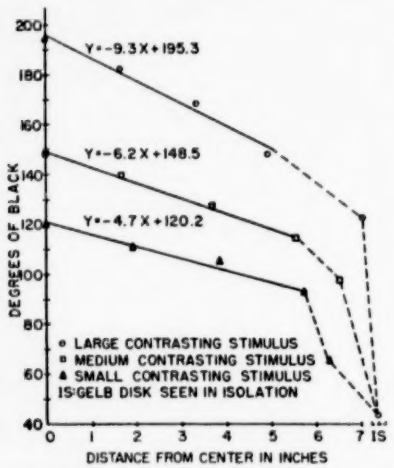

FIG. 3. Theoretical curves and empirical averages of observations made by three *S*s: JMc, DV, and JM.

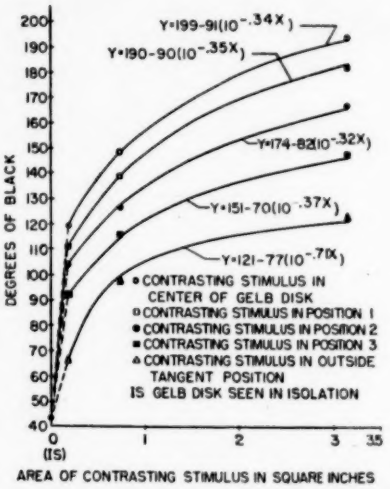

FIG. 4. Theoretical curves and empirical averages of observations made by three *S*s: JMc, DV, and JM.

show the results as functions of the size of the contrasting stimuli and their positions on the surface of the disk. An inspection of this figure shows that the darkening of the Gelb disk is a function of the size of the contrasting stimulus.

In plotting the data of the size of contrasting stimulus versus appearance of the Gelb disk in degrees of black, the values for the absence of, or zero area of, contrasting stimulus were not included in the curves which are fitted to the data. As can be seen in Fig. 4, zero area of contrasting stimulus represents a mathematical as well as an experimental discontinuity. Hence, the values in degrees of black are joined by a dotted line to the curves for the three sizes of contrasting stimuli.

The lightness of the disk as seen by *S*s differed greatly. Two *S*s, DV and JMc, perceived the disk at about the same degree of lightness, but JF and JM differed from each other and from the other two *S*s. (It should be remembered that JF had to be tested in a dimmer illumination than the other three *S*s.) JM perceived the disk as lighter than did the other *S*s when tested with the Gelb disk in isolation. The contrasting stimuli darkened the disk for JM to a lesser extent than for any of the other *S*s. JF perceived the Gelb disk in isolation as darker than did any of the other *S*s. The contrasting stimuli produced a greater darkening of the disk for JF than for JM, but the greatest darkening effect was obtained in the data of JMc and DV.

All *S*s were alike in a second respect, however. They had a tendency to be consistent during the same session, but varied from session to session (see Table 3).

The one exception to the general conclusion that all *S*s perceived a darker Gelb disk as the size of the contrasting stimulus was increased is the individual data of JM.

The values of JM where the small and medium contrasting stimuli were used are very close together on the left side of the disk. The average distance, excluding the outside position, is less than 1°. At Position 3, the small contrasting stimulus turned the disk darker than the medium one at the same position. The difference is 0.5°. When, however, the measurements from the two sides of the disk are averaged, this one discrepancy is eliminated.

Each of the five positions of the contrasting stimuli used in the experiment was found to affect the lightness of the Gelb disk. The greatest dark-

TABLE 3

JUDGMENTS GIVEN BY FOUR *S*S FOR DIFFERENT CONDITIONS OF THE GELB DISK

Position of Contrasting Stimulus[a]	Size of Contrasting Stimuli											
	Small				Medium				Large			
	JMc	DV	JF	JM	JMc	DV	JF	JM	JMc	DV	JF	JM
C	124.9	132.4	179.9	102.0	174.0	159.3	201.8	111.3	227.7	208.9	236.5	144.2
1	117.9	125.3	172.0	89.0	163.3	146.4	193.4	105.3	210.0	202.9	224.8	130.7
2	115.8	114.4	167.5	84.0	148.0	137.8	185.0	93.8	195.3	186.5	218.8	120.0
3	97.8	101.8	161.5	77.3	131.3	130.8	178.0	81.3	156.5	176.8	199.8	108.7
O-T	67.9	76.8	152.9	53.0	114.5	107.4	162.8	69.8	138.2	129.8	181.5	98.1

Note.—Figures represent the average of four trials in degrees of black on the measuring disk.
[a] The symbols used for identification of position have the following meanings: C, center; O-T, outside-tangent. The numbers are intermediate positions with measurements as given in Table 2.

ening effect was produced by the contrasting stimulus in the center position. Diminishing effectiveness was produced as the contrasting stimulus was moved away from the center. The least effective position was on the outside but tangent to the edge of the disk (Fig. 4). The data of JM yield one exception to the above generalization. In his case, both the center and the R1 positions with the medium-sized contrasting stimulus gave the same match. When the measurements for the right side of the disk are averaged with the measurements of the left side, however, the two positions are different and in the expected direction.

Here again it is to be noted that the outside-tangent represents both an experimental and a mathematical discontinuity, hence the extrapolation to this position is shown by a dotted line. Inspection of Fig. 3 shows that with the outside-tangent position excluded, the data fit a straight line. JF was not included, since this *S* required different conditions of illumination.

The next step taken to show the effects of position of the contrasting stimuli on the Gelb disk is shown in Fig. 5, where the three sizes of the contrasting stimuli are confounded. Curve A is the theoretical curve for the average values of JMc, DV, and JM, and for the four positions of center, 1, 2, and 3. The empirical averages for the four *S*s are plotted on the figure, and show that Curve A is closely approximated by the empirical values. There is only one discrepant average value, JMc. Since the curves of the four *S*s are found at different levels of degrees of black, an equalization of the curves was necessary to allow a direct comparison of their slopes. Hence the midpoint of each curve was plotted at a zero-value of slope units.

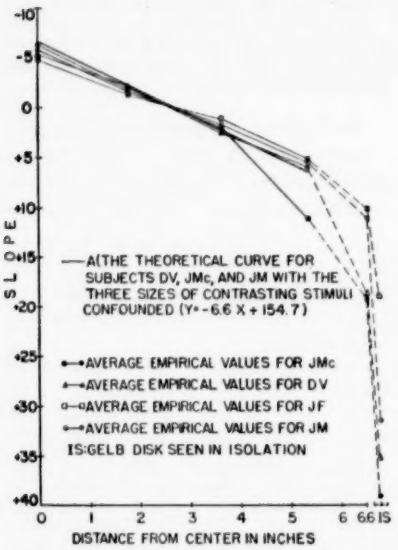

FIG. 5. Slopes of the individual curves for four positions of contrasting stimuli compared with A.

DISCUSSION

The results of this experiment make quite clear that the Gelb effect is essentially a result of contrast and hence subject to the usual laws of lightness-contrast. This is a confirmation of Hypothesis 1. A description of the changes in the Gelb disk is in good accord with both the inferences drawn from the data and with contrast phenomena in general.

The fact that the luminance of the field surrounding the measuring disk was different from that of the Gelb disk is fully appreciated. If the luminance of the field around the measuring disk is equal to that of the field around the Gelb disk, the settings of the two will be equal, within the limits of experimental error and no measurement of the Gelb effect is possible. The absolute values (degrees of black) reported apply

only to the conditions of the luminances of the two fields surrounding the Gelb and measuring disks employed in this study. Different levels would be expected to yield another set of absolute values. But the shifts in lightness of the Gelb disk found here as a function of size and position of the contrasting stimulus should be found with different luminances, at least over a very wide range. Our interest is, therefore, directed not only to the actual lightness or darkness of the Gelb disk under any given condition of viewing, but just as much to the changes in the Gelb effect occasioned by the variation in conditions employed in this study.

The darkening effect induced by the contrasting stimulus was not uniform over the entire surface of the Gelb disk. The greatest darkening occurred in the region of the contrasting stimulus, and diminished with distance from it. It follows from this observation that the maximal darkening of the disk as a whole is obtained when the contrasting stimulus is in the center of the disk, an expectation borne out by the data. However, in reporting the effect of the position of the contrasting stimulus in darkening the Gelb disk, the outside-tangent position was excluded from consideration. Plots of the data indicated that a discontinuity existed between the outside-tangent and the No. 3 position. The reasons for the discontinuity are not clear. Probably significant, is the fact that the contrasting stimulus in the outside-tangent position was outside the cone of light focused on the Gelb disk, and hence was less bright than the contrasting stimulus in any other position. A second factor that could be important in producing the discontinuity in the data, is the variable of configuration. Its effectiveness in inducing contrast effects has been well recognized (Hartmann, 1935; Jenkins, 1930; Mikesell & Bentley, 1930).

A second discontinuity was found in the parameter of size of the contrasting stimulus. The discontinuity was obtained between the smallest size of the contrasting stimulus used, .20 sq. in., and the absence of a stimulus (Fig. 4). When the absence of a contrasting stimulus is excluded from consideration, the relationship between the size of the contrasting stimulus and the changes in the Gelb disk is expressed by a growth function. These results are equivocal in meaning, but are compatible with a contrast interpretation of the Gelb phenomenon.

Hypothesis 2 was that the lightness of the Gelb disk is a function of *all* reflectances in the field of view. This hypothesis was also confirmed. The smallest contrasting stimulus used, a circle of .5-in. diameter describing a visual angle of approximately 10 min., produced a dependable and measureable darkening effect on the Gelb disk. Osgood (1953) and Woodworth and Schlosberg (1954) have objected to a contrast explanation of the Gelb phenomenon on the grounds that the contrasting stimuli used are too small to induce contrast effects. Yet, the quantitative results of the experiment clearly indicate that the Gelb phenomenon is based on an induced contrast effect.

A final indication of the contrast nature of the Gelb phenomenon is found in the confirmation of Hypothesis 3, which was that the effect of a reflectance will increase as the area of the contrasting stimulus is increased, and as the reflectance area is brought nearer the center of the disk. The verification of these two points constitutes the core of the present study. The results of the experiment confirm all inferences from the hypothesis and thereby systematically related the Gelb phenomenon to contrast effects.

Summary

A spot of light was focused to coincide with the boundaries of a black rotating disk, thereby producing an appearance of grayish-white. This effect was studied by varying two factors. White contrasting stimuli of the same shape but of three different sizes were successively suspended in front of the disk. The positions of those contrasting stimuli were varied from the center of the disk to a location on the outside and tangent to the disk. Three intermediate positions were also

included in the design. The perceived appearance of the Gelb disk was recorded by a measuring disk adjustable by the *E* from 360° white to 360° black.

It was found that increase in size of the contrasting stimuli darkened the appearance of the disk. The greatest effect was obtained when the stimulus was in the center of the disk. The darkening effect diminished linearly as the stimulus was moved to the inside-tangent position. The least darkening effect occurred with the stimulus in the outside-tangent position. Equations for the data were determined that were based on the size and position of the contrasting stimuli as well as the apparent lightness of the disk.

The results of the experiment were discussed in light of the criticism that the area of the smallest stimulus used was too small to produce contrast-effects obtained. The quantitative data showed that the Gelb phenomenon is a contrast-effect governed by the usual laws of lightness-contrast.

REFERENCES

DIAMOND, A. L. Foveal simultaneous contrast as a function of inducing field area. *J. exp. Psychol.*, 1955, **47**, 144–152.

GELB, A. Die "Farbenkonstanz" der Sehdinge. In A. Bethe, G. v. Bergmann, G. Embden, & A. Ellinger (Eds.), *Handbuch der Normalen und Pathologischen Physiologie*. Vol. 12. Berlin: Julius Springer, 1929. Part I, pp. 594–678.

HARTMANN, G. W. *Gestalt psychology*. New York: Ronald Press, 1935.

HELSON, H. Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. exp. Psychol.*, 1938, **23**, 439–476.

HELSON, H. Some factors and implications of color constancy. *J. opt. Soc. Amer.*, 1943, **33**, 555–567.

HELSON, H. Adaptation-level as frame of reference for the prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HENNEMAN, R. H. A photometric study of the perception of object color. *Arch. Psychol., N. Y.*, 1935, No. 179.

JENKINS, J. A. Perceptual determinants in plane designs. *J. exp. Psychol.*, 1930, **13**, 24–46.

JUDD, D. B. The definition of black and white. *Amer. J. Psychol.*, 1941, **54**, 289–294.

KARDOS, L. Ding und Schatten. *Zsch. f. Psychol.*, 1934, Suppl., 23:184.

KATZ, D. *The world of colour*. London: Kegan Paul, 1935.

KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.

MIKESELL, W. H., & BENTLEY, M. Configuration and contrast. *J. exp. Psychol.*, 1930, **13**, 1–23.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Henry Holt, 1954.

(Received May 5, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

INTERSENSORY COMPARISONS OF TEMPORAL JUDGMENTS [1]

SANFORD GOLDSTONE, WILLIAM K. BOARDMAN, AND WILLIAM T. LHAMON

Baylor University College of Medicine and Houston State Psychiatric Institute

Comparisons of absolute judgments of stimuli presented to different senses are made difficult by the inappropriateness of most psychological dimensions for multisensory judgments. Concepts of time, however, may be considered similar regardless of the sense mode involved.

This report describes studies of the relationship between auditory and visual judgments of time beyond the effects of accessory and ambient stimuli (Hirsh, Bilger, & Deatherage, 1956; London, 1954; Ryan, 1940).

It has been shown that absolute judgments of a standard unit of time, 1 sec., using auditory durations, change with varied anchor conditions (Goldstone, Lhamon, & Boardman, 1957), developmental sequence (Smythe & Goldstone, 1957; Goldstone, Boardman, & Lhamon, 1958b), toxic states (Boardman, Goldstone, & Lhamon, 1957; Goldstone, Boardman, & Lhamon, 1958a), and psychopathology (Lhamon & Goldstone, 1956; Weinstein, Goldstone, & Boardman, 1958). The present study was carried out to investigate audio-visual differences in absolute judgments of *S*'s concept of one clock second derived from judgments of lights and sounds under several experimental conditions: (*a*) Successive auditory and visual judgments were obtained in the same experimental setting in order to study the effects of prior experience with one sense mode upon temporal judgment with another. (*b*) Long and short stimulus anchors were employed in the same experimental setting in order to study the differential effects of recent and remote anchors upon auditory and visual judgments of the apparent duration of one clock second. (*c*) Auditory and visual signals were presented simultaneously, with varied intensity and instructions to focus upon either the lights or the sounds, in order to explore sensory dominance and stimulus effectiveness.

[1] Supported in part by grant M-1121, U.S.P.H.S. and by the U.S.A.F. under contract AF18(603)-79, monitored by the AFOSR of the Air Research and Development Command. The authors are indebted to Joyce Goldfarb and Lynn Koehler, Research Assistants, who helped implement this project.

GENERAL METHOD

Apparatus.—An electronic timer activated auditory and visual signals for durations of .01 to 9.99 sec. An audio-oscillator and headphones provided a tone of 725 cps at 70 db (*re* .0002 dyne/cm.^{2}). The light was a glow modulator tube which illuminated a 5-mm. circular test patch mounted in a black box at 1° visual angle. The source was pulsed at 450 cps with a light-dark ratio of 50:50, and appeared as a white light of approximately 6 apparent ft.-candles.

Method.—The *S*s were college students. Age and sex were not controlled (Smythe & Goldstone, 1957). Durations were presented in accordance with a modified method of limits technique using either a geometric 10% stimulus series or an arithmetic .10 sec. series and *S* responded to each duration as "more" or "less" indicating an absolute judgment of more or less than 1 sec. No standard for comparison was provided. The *S*s received 20 alternating ascending and descending runs, each ascending run terminating upon three consecutive reports of "more" and each descending run terminating with three "lesses."

The percentages of reports of "less" for all

TABLE 1

THE MEDIAN SECOND ESTIMATION POINT (SEP) IN SECONDS UNDER ALL SENSORY CONDITIONS (EXP. I)

Group	Modality	*N*	*Mdn.*	*Q*
A	Auditory	41	.59	.24
V	Visual	40	1.02	.36
AV	Auditory	21	.69	.26
	Visual		1.00	.40
VA	Visual	20	1.21	.18
	Auditory		.76	.20

durations of each experimental condition were plotted on arithmetic probability paper and a straight line visually fitted. The Second Estimation Point (SEP), or duration which *S* reported more or less than 1 sec. 50% of the time, was derived from these plots. The SEP data were treated with Mann-Whitney and Wilcoxon tests.[2]

Three experiments were performed. There was no overlap of *Ss* from experiment to experiment.

EXPERIMENT I

Procedure.—The *Ss* ($N = 122$) were divided into four groups: 41 auditory only (A), 40 visual only (V), 21 auditory followed by visual (AV), 20 visual followed by auditory (VA). All *Ss* received 20 runs starting at 1.00 sec. with a 10% geometric series. If *S* reported "more" to the initial stimulus, the first run was descending; if *S* reported "less," the first run was ascending. A 1-min. rest separated the auditory and visual series for Group AV and VA. Group A only judged sounds and Group V only judged lights.

Results.—Table 1 summarizes the median SEP values for Groups A and V and for both sequences of Groups AV and VA. The visual SEP is greater than the auditory for both the intergroup and the intragroup comparisons ($P < .01$). The only exception to this is when the visual SEP's of AV are compared with the auditory SEP's of VA. Inspection suggests that the final auditory SEP's were "pulled" up by the initial visual series, and the final visual SEP's were "pulled" down by the initial auditory series. However, this intersensory transfer effect is not reliably demonstrated in these data.

[2] An appropriate log transformation and analysis of variance were done, but since the results conformed to those obtained with the nonparametric analysis, the clearest method is reported.

EXPERIMENT II

Procedure.—The *Ss* ($N = 241$) were divided as follows: 40 visual short anchor followed by visual long anchor (VSL), 40 visual long anchor followed by visual short anchor (VLS), 40 visual neutral anchor (VN), 40 auditory short anchor followed by auditory long anchor (ASL), 40 auditory long anchor followed by auditory short anchor (ALS), 41 auditory neutral anchor (AN).

Groups VSL and ASL were tested with a .10-sec. arithmetic series starting at .10 sec. (short anchor). Following a 1-min. rest, *S* received a series of runs starting at 2.00 sec. (long anchor). Groups VLS and ALS were similarly tested but with the first series starting at 2.00 sec. (long anchor) and the final series starting at .10 sec. (short anchor). Groups VN and AN were tested starting at 1.00 sec. (neutral anchor).

Results.—Tables 2 and 3 summarize the median SEP's for the visual and auditory groups under all anchor con-

TABLE 2

MEDIAN VISUAL SECOND ESTIMATION POINT (SEP) IN SECONDS UNDER ALL ANCHOR CONDITIONS (EXP. II)

Group	Anchor	*N*	*Mdn.*	*Q*
VN	Neutral	40	.98	.31
VSL	Short	40	.66	.22
	Long		.77	.24
VLS	Long	40	1.28	.42
	Short		1.02	.34

TABLE 3

MEDIAN AUDITORY SECOND ESTIMATION POINT (SEP) IN SECONDS UNDER ALL ANCHOR CONDITIONS (EXP. II)

Group	Anchor	*N*	*Mdn.*	*Q*
AN	Neutral	41	.56	.20
ASL	Short	40	.40	.11
	Long		.55	.19
ALS	Long	40	.71	.33
	Short		.55	.31

ditions. For both sense modes, the initial short anchor accompanied the shortest SEP's, the initial long anchor accompanied the longest SEP's, and the neutral anchor group was in the middle. The anchor stimuli exerted a pulling effect on the judgmental transition zone in the direction of the anchor magnitude for auditory and visual judgments. The VSL short anchor series had significantly smaller SEP's than either VLS long anchor or VN ($P < .001$). The VLS long anchor series was greater than VN ($P < .05$). The ASL short anchor series was smaller than either ALS long anchor ($P < .001$) or AN ($P < .01$). The ALS long anchor group was not significantly greater than AN.

Remote auditory anchors influenced judgment in a new anchor context. The SEP's obtained with the long anchor preceded by the short anchor were not different from those obtained with the short anchor preceded by the long one. An interaction of anchor influences is suggested.

Remote visual anchors also influenced judgment in a new anchor context but in a slightly different fashion. It appears that the remote anchor has a dominant influence over subsequent judgments under new anchor conditions. Past visual experience with the long anchor had a greater pulling effect than a contemporary short visual anchor, and past experience with the short anchor influenced judgment more than a more recent long anchor. Although VLS long anchor had significantly higher SEP's than VLS short anchor, and VSL long anchor was more than VSL short anchor ($P < .001$), VSL short anchor was significantly less than VLS short anchor, and VLS long anchor was higher than VSL long anchor ($P < .001$) emphasizing the dominance of the initial anchor for the visual mode, whereas more simple interactions of remote and recent anchor effects appear with auditory judgments.

Figure 1 compares median auditory and visual SEP's under all initial and final anchors. As noted in Exp. I, the auditory groups obtained smaller SEP's than the visual groups under each initial anchor condition (ASL < VSL, $P < .01$; AN < VN and ALS < VLS, $P < .001$). This consistent auditory-visual difference makes it difficult to evaluate possible intersensory differences in anchor effects. The greater slope of the Anchor-SEP plots in Fig. 1 for the visual groups suggests that this mode is more affected by stimulus anchors.

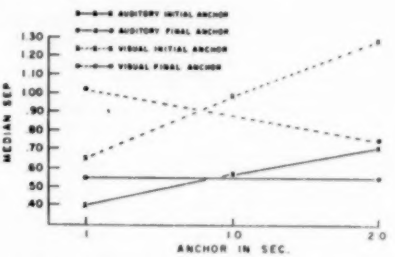

FIG. 1. Median SEP-Anchor plots for auditory and visual judgments.

EXPERIMENT III

Procedure.—Preliminary study indicated that *Ss* consistently rated a light of 6 apparent ft.-candles as more intense than a 53 db tone of 725 cps; and a tone of 78 db was rated as more intense than a light of 4 apparent ft.-candles. None of the intensities was considered unpleasant. Forty *Ss* were divided into two equal groups and were given sounds and lights simultaneously, starting at 1.00 sec. using a .10-sec. arithmetic series. The sound was more intense for one group (Av), and the light for the other (Va). After a 1-min. rest, half of Group Av was asked to focus attention on the sound while receiving both signals, and the other half was asked to focus upon the light. Half of Group Va was also asked to focus on either the light or the sound.

Results.—Table 4 summarizes the median SEP's obtained with simultaneous auditory-visual stimulation. The SEP's obtained when the sound was more intense were smaller ($P < .001$) than the SEP's obtained when the light was more intense. When the sound was more intense the SEP's were approximately the same as those obtained with auditory alone; with the light more intense, the SEP's approximated those of visual alone.

Differential focus of attention played no significant role in SEP determination. When the light was more intense *Ss* obtained longer SEP's regardless of whether they focused upon light or sound and when the sound was more intense *Ss* obtained shorter SEP's regardless of focus.

TABLE 4

MEDIAN SECOND ESTIMATION POINT (SEP) IN SECONDS WITH VARIED INTENSITY, VARIED ATTENTION AND SIMULTANEOUS STIMULATION (EXP. III)

Group	Attention	*N*	*Mdn.*	*Q*
Av	No focus	20	.50	.19
	Sound focus	10	.59	.22
	Light focus	10	.64	.15
Va	No focus	20	.99	.28
	Sound focus	10	.86	.22
	Light focus	10	1.04	.43

DISCUSSION

No obvious explanation accounts for the longer visual second. The anchor effects indicate that this audio-visual difference is not simply a function of greater accuracy of visual conceptual time judgments. If the magnitude of the long anchor were increased, the auditory judgments would become more accurate and the visual judgments would depart further from the physical standard. It is more probable that there is a fundamental difference between auditory and visual time perception such that more visual clock time is judged necessary to fulfill a basic temporal concept. Visual durations that were the same as auditory durations were judged shorter.

It is generally recognized that absolute judgments of magnitude and intensity are relative matters determined largely by *S*'s frame of reference, and are influenced not only by past experience but by the range of magnitudes and extreme stimuli in a given context (Helson, 1947). Although subjective scale shifting due to anchor effects has been demonstrated along many psychological dimensions (Guilford, 1954) employing a variety of sense modes, direct comparisons of intersensory differences are difficult to accomplish because most dimensions are not analogous across the senses. Absolute judgments of standard concepts of time permit such transmodal comparisons and this study demonstrated subjective scale shifting in the direction of the anchor magnitude for both auditory and visual judgments. It was suggested that the visual judgments were more influenced by anchors.

It was demonstrated that past contact with durations presented to one mode influenced subsequent judgments with the other. Also remote anchor stimuli influenced contemporary time judgments,

with past anchors more potent in affecting visual judgments. Although interaction is demonstrated when past-present sequences are considered, simultaneous audio-visual presentation may involve an all-or-none judgment with stimulus intensity a primary factor. When the sound was judged more intense than the light, the judgments approximated those obtained with auditory alone, and when the light was more intense the judgments approximated those obtained with visual alone. The focus of attention on either the light or the sound did not influence judgment.

Hirsh et al. (1956) found no audio-visual differences in the reproduction of 1-sec. durations. However, his experiment differed from ours in that kinesthetic cues were involved in the reproduction method, and the standard for comparison was an external stimulus, not an internal concept.

It is not possible to generalize our findings to all kinds of temporal judgments and all durations. Nor is it valid to make general comparative statements concerning visual and auditory judgment. Absolute judgments of concepts of duration apparently reflect flexible intra-organismic norms and intersensory differences. It would be fruitful to extend this work to include intersensory comparisons of temporal judgments using methods other than passive estimation and to include durations other than 1 sec.

Summary

Audio-visual comparisons were made of absolute judgments of a standard unit of time, 1 sec. The *Ss* judged the duration of sounds and lights as more or less than their concept of one clock second. Four hundred and three *Ss* were divided into 12 groups: 41 auditory only, 40 visual only, 21 auditory followed by visual, 20 visual followed by auditory, 40 visual short anchor followed by visual long anchor, 40 visual long anchor followed by visual short anchor, 40 visual neutral anchor, 40 auditory short anchor followed by auditory long anchor, 40 auditory long anchor followed by auditory short anchor, 41 auditory neutral anchor, 20 simultaneous audio-visual with auditory more intense, 20 simultaneous audio-visual with visual more intense. The following findings were reported and discussed:

1. The visual "second" was consistently longer than the auditory "second."

2. Long and short anchor stimuli pulled the transition zone of the *Ss*' subjective "second" in the direction of the anchor magnitude for both modes. Visual judgments were more affected by the anchors.

3. Past experience with long and short anchors influenced contemporary temporal judgments, with the residual effect of a past anchor being more potent for the visual mode.

4. Stimulus effectiveness and sensory dominance appeared to be a function of stimulus intensity. Under conditions of simultaneous audio-visual stimulation, the judgments approximated auditory alone when auditory was more intense; when visual was more intense the judgments approximated visual alone.

5. Stimulus intensity was the differentiating factor even when attention was varied through instructions to focus on either the light or the sound.

References

Boardman, W. K., Goldstone, S., & Lhamon, W. T. Effects of lysergic acid diethylamide (LSD) on the time sense of normals. *Arch. Neurol. and Psychiat.*, 1957, **78**, 321–324.

Goldstone, S., Boardman, W. K., & Lhamon, W. T. Effect of quinal barbitone, dextroamphetamine, and placebo on apparent time. *Brit. J. Psychol.*, 1958, **49**, 324–328. (a)

Goldstone, S., Boardman, W. K., & Lhamon, W. T. Kinesthetic cues in the development of time concepts. *J. genet. Psychol.*, 1958, **93**, 185–190. (b)

Goldstone, S., Lhamon, W. T., & Boardman, W. K. The time sense: Anchor effects and apparent duration. *J. Psychol.*, 1957, **44**, 145–153.

Guilford, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954. Pp. 312–314.

Helson, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

Hirsh, I. J., Bilger, R. C., & Deatherage, B. H. The effect of auditory and visual

background on apparent duration. *Amer. J. Psychol.*, 1956, **69**, 561–574.

LHAMON, W. T., & GOLDSTONE, S. The time sense: Estimation of one second durations by schizophrenic patients. *Arch. Neurol. and Psychiat.*, 1956, **76**, 625–629.

LONDON, I. D. Research on sensory interaction in the Soviet Union. *Psychol. Bull.*, 1954, **51**, 531–568.

RYAN, T. A. Interrelations of the sensory systems in perception. *Psychol. Bull.*, 1940, **37**, 659–698.

SMYTHE, E. J., & GOLDSTONE, S. The time sense: A normative genetic study of the development of time perception. *Percept. mot. Skills*, 1957, **7**, 49–59.

WEINSTEIN, A. D., GOLDSTONE, S., & BOARDMAN, W. K. The effect of recent and remote frames of reference on temporal judgments of schizophrenic patients. *J. abnorm. soc. Psychol.*, 1958, **57**, 241–244.

(Received May 6, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

PERFORMANCE IN EYELID CONDITIONING AS A FUNCTION OF UCS DURATION [1]

W. N. RUNQUIST [2] AND K. W. SPENCE

State University of Iowa

A recent series of investigations (Spence, 1957; Spence, Haggard, & Ross, 1958) has shown that when care was taken to equate the drive level (*D*) of *S*s, performance in classical aversive conditioning was directly related to the intensity of the UCS on reinforcing trials. This finding was interpreted as reflecting a difference in the learning factor (*H*) and hence as supporting a reinforcement type of learning theory so far as aversive conditioning is concerned. While it was indicated that our preference was to confine interpretation to the more general conception that *H* is some unspecified kind of function of the intensity of the UCS, one can advance the more specific drive-stimulus reduction hypothesis of Hull (1943) and Miller and Dollard (1941). According to this view, the cessation of a strong UCS would provide greater reinforcement and hence a greater increment of *H* per trial than termination of a weak UCS.

One implication of this drive-stimulus reduction hypothesis is that the greater the time interval between the occurrence of the response and the reinforcing event, the poorer should be the conditioning performance. In most classical aversive conditioning experiments the noxious UCS is presented for only a brief period with the consequence that the response, which occurs with the onset of the UCS, is followed almost immediately by the reinforcing event, the termination of the UCS. Delay of the reinforcement could be produced by increasing the duration of the UCS.

The present studies were designed to test the implications of this drive-stimulus reduction theory of reinforcement by comparing eyelid conditioning performance with different durations of the UCS.

METHOD

Subjects and design.—The *S*s were 160 introductory psychology students, assigned to one of two groups in order of appearance at the laboratory. One group received a 50-msec. air puff as a UCS, the other group a 1000-msec. air puff. Four independent experiments were run which deviated slightly in procedure. In each of these experiments there were 12 males and 8 females in a group. An additional 8 *S*s were eliminated as voluntary responders (Spence & Taylor, 1951). The first two studies differed in the location and intensity of the UCS. In the first experiment, the UCS was a 1-lb. puff delivered to the cornea of the right eye. In the second experiment, the UCS was an 8.3-lb. air puff delivered to the side of *S*'s head 2 cm. from the cornea of the right eye. The purpose of the second study was to prevent *S* from reducing the noxiousness of the stimulus by closing his eye. The second two studies were replications of the first two with one exception. In the first two studies the CS duration was 550 msec. for both groups. In the second two studies the CS overlapped the UCS termination by 50 msec. for both UCS durations.

Apparatus.—The equipment for delivering the UCS and the CS, the increase in bright-

[1] This study was carried out as part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802, Project NR 154–107 between the State University of Iowa and the Office of Naval Research. Acknowledgment is made to John Hunter and David Stubbs who served as research assistants in the above investigation.

[2] Now at Hobart College.

ness of a 6 cm. circular disc, was the same as that used in previous studies (Spence, Farber, & Taylor, 1954). The UCS onset occurred 500 msec. following the onset of the CS for all groups. The duration of the UCS was controlled in each experiment by means of a 110-v. AC-operated solenoid valve controlled by an electronic timer. Changes in the strength of the air puff through its duration were limited to approximately .12 lb. by means of a storage tank in the air line immediately in front of the solenoid valve.

Procedure.—All *S*s were given 100 conditioning trials under the same procedure. There were initially three trials of the CS alone followed by one trial with the UCS alone. Each trial was preceded by a verbal ready signal 2–4 sec. before the onset of the CS. The intertrial interval was 15, 20, or 25 sec. in a prearranged order. A CR was recorded whenever a deflection in the record of more than 1 mm. occurred 200–500 msec. following the onset of the CS.

Results

The curves of the acquisition of the CR based on the percentage of CR's occurring in successive blocks of 10 trials are presented in Fig. 1. All groups having the 50-msec. UCS have been combined and all groups having the 1000-msec. UCS have been combined. The results show a slight but consistent superiority in performance for the *S*s conditioned with the shorter UCS. An analysis of variance showed that of the four main effects only Location of air puff (eye vs. face) and Sex were significant sources of variation, while Sex × Overlap-nonoverlap was the only significant interaction. In this analysis UCS duration fell short of significance ($F = 1.84$ with 1 and 145 df). Since the four experiments were not conducted simultaneously but in order, the unambiguous interpretation of these differences is not possible. Furthermore, the fact that location of puff made a difference cannot be interpreted because there was also a difference in the intensity of the UCS. In determining this intensity for the "face" condition, an explicit attempt was made to make this UCS strong enough to produce approximately the same performance level as that produced by the 1-lb. puff to the eye. The significant F indicates that this was not successful, as performance was slightly lower in the groups having the 8-lb. puff to the face. Despite the fact that there were differences in performance level in the four experiments, it is felt that the lack of interactions with UCS duration (all F's less than 1.00) justifies the pooling of the data from the four studies.

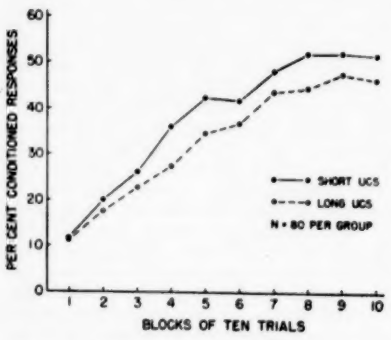

FIG. 1. Percentage of CR's in blocks of 10 trials for all 50-msec. groups combined and all 1000-msec. groups combined.

Although the F for UCS duration was not significant, a Mann-Whitney test on these combined groups gave a value of 1.68 which is significant at the .05 level for a one-tailed test.[3] Table 1 presents further data supporting a conclusion that the shorter UCS produces a slightly higher level of performance. If men and women are considered separately, eight independent comparisons between the

[3] It has been found in several studies in this laboratory that the Mann-Whitney test is more sensitive than the t test. The value of t obtained with these data was 1.29.

TABLE 1

MEAN CR's BY INDIVIDUAL GROUPS

CS	Sex	UCS	50 Msec.	1000 Msec.	Diff.
Nonoverlap	M	Eye	42.5	40.3	2.2
Nonoverlap	M	Face	30.7	28.7	2.0
Overlap	M	Eye	29.2	26.7	2.5
Overlap	M	Face	24.6	22.4	2.2
Nonoverlap	F	Eye	51.6	39.3	12.3
Nonoverlap	F	Face	39.9	30.8	9.1
Overlap	F	Eye	62.5	53.9	8.6
Overlap	F	Face	39.5	30.9	8.6

50-msec. and 100-msec. conditions are provided. As may be seen from the last column of the table, these eight differences all favor the short UCS condition, with the size of the difference being extremely consistent within sexes, but somewhat larger for the women. Although no one of these differences is statistically significant, the probability of obtaining all eight differences in the same direction under the null hypothesis is .004.

DISCUSSION

It seems possible to conclude from these data that increasing the duration of the UCS in classical aversive conditioning lowers the performance level in this type of aversive conditioning, although the effect within the range of values studied here is small. Nevertheless, the results do provide some support for the drive-stimulus reduction hypothesis.

While the deduction concerning the effect of UCS variation is straightforward, the experiment is possibly confounded by factors which would tend to reduce the effect of delaying reinforcement. Thus it is not unreasonable to assume that the drive level (D) produced by a relatively long UCS duration is greater than that produced by a short duration. If this were the case, it would tend to reduce any difference in performance due to differential delay. A study by Bitterman, Roth, & Krauskopf (1952) has attempted to control this effect by simultaneously conditioning a GSR to two stimuli, one with a short shock and the other with a long shock. Thus, each S receives both short and long UCS. There was no difference in conditioning to the stimulus paired with the short shock as opposed to the stimulus paired with the long shock; however, the two stimuli were so similar in this study that it is doubtful whether any differentiation took place.[4]

Even if the over-all drive level were controlled in this way there could still be differential drive reduction, with greater drive reduction occurring with the cessation of the long UCS. If H is related to the amount of drive reduction, the greater H produced with a long UCS would counteract the performance decrement produced by the delay. It appears impossible to isolate these factors completely.

In spite of the fact, then, that these two possible factors were working against obtaining a higher level of performance with a short UCS duration, this result was obtained. Even though the difference is quite small and only on the borderline of statistical significance, it appears to be replicable, thus supporting the drive-stimulus reduction hypothesis. This finding is also consistent with those of Sullivan (1950) who found a higher level of GSR conditioning with a short UCS than with a long UCS, although his data are not completely unequivocal.

SUMMARY

An experiment was performed to test an implication of the drive-stimulus reduction hypothesis, that the greater the time interval between the response and the reinforcing event, the poorer should be the conditioning performance. Delay was produced by increasing the duration of the UCS. Two groups of Ss in each of four experiments were conditioned with either a 50-msec. or 1000-

[4] The fact that these investigators eliminated all Ss who could not verbalize the distinction between stimuli does not insure that a *conditioned* discrimination was established in terms of the GSR response.

msec. air puff as a UCS. The four experiments differed in minor procedural variations.

When all experiments are combined, the 50-msec. group showed significantly higher performance level over the last 50 trials. In eight out of eight comparisons the short UCS group were superior, although the differences were quite small. The results were interpreted as tending to support the drive-stimulus reduction hypothesis, especially when considered in relation to possible factors operating in this situation which would tend to work in a direction opposite to that of the UCS duration variable.

REFERENCES

BITTERMAN, M. E., ROTH, P., & KRAUSKOPF, J. Effect of duration of the unconditioned stimulus on conditioning and extinction. *Amer. J. Psychol.*, 1952, **65**, 257–262.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1957.

SPENCE, K. W., FARBER, I. E., & TAYLOR, E. Relation of electric shock and anxiety level to performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

SPENCE, K. W., HAGGARD, D., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SULLIVAN, J. J. Some factors affecting the conditioning of the galvanic skin response. Unpublished Ph.D. dissertation, State Univer. of Iowa, 1950.

(Received May 7, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

THE RELATION OF ACHIEVEMENT IN PROBLEM SOLVING TO RATE AND KIND OF HYPOTHESES PRODUCED[1]

RAMON J. RHINE

System Development Corporation

One important aspect of problem solving is the formation by *S*s of ideas or hypotheses about a problem's solution. Evidence, reported by Sargent (1940), suggests that both the rate of hypothesis production and the type of hypothesis produced might be significant determinants of problem-solving achievement. One form of hypothesis behavior, "use of analysis," refers to analyzing anagrams by searching for letter combinations, such as prefixes or suffixes, that seem to make sense in terms of past experience with words. This mode of behavior may be contrasted with a whole approach in which *S* attains solution after "reading letters and looking at them with no attempt to form syllables or combinations" (Sargent, 1940, p. 19). The contrast between use of analysis and the whole approach is reminiscent of Heidbreder's (1924) distinction between "participant" and "spectator" behavior. Sargent's data tentatively suggest that achievement with anagram problems depends in part upon the type of hypothesis—letter combination or whole words—which *S* tends to use.

The present study is an investigation of the relation to problem-solving achievement of rate and kind of hypotheses produced with anagram and spatial problems. Hypotheses for anagram problems are readily separated into two classes. With scores on spatial problems available, it is possible to determine whether the use of a given class of anagram hypotheses is related to achievement on a different type of problem. Such a relationship would provide evidence for the occurrence, as a factor in problem solving, of an ability to select the more effective of alternative classes of hypotheses.

METHOD

Materials and subjects.—Of 34 male undergraduates, all were given six paper-and-pencil, spatial problems to solve, and 32 returned to take six verbal problems. Three of the spatial problems are shown in a report by Sweeney (1953, pp. 71–72, Problems B-3, B-4, and B-5), and two further spatial problems are given by Taylor and McNemar (1955, p. 16, Problems 2 and 3). The sixth problem is the familiar pyramid of 10 pennies, in which the pyramid is to be turned upside down by moving only three pennies. The verbal problems were all anagrams, each being presented along with the category into which its solution falls, as follows: famous American—*eeolorstv* (Roosevelt); plant life—*a low life cur* (Cauliflower); famous American—*eiycklmn* (McKinley); State—*dash red lion* (Rhode Island); plant life—*a sugar sap* (asparagus); plant life—*i hate cork* (artichoke). Both types of problems were selected on the basis of preliminary information indicating that about 50% would be solved if a 7-min. time limit was imposed on each problem. With each spatial problem, for which the answer could be shown by a drawing, a set of cards bearing small reproductions of the problem figure were made avail-

[1] This experiment was one of a series of studies in problem solving done under Project NR 150-149 and supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Work on the contract was under the general direction of Donald W. Taylor. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government.

able, and a set of blank cards was given to *S* as each anagram was administered.

Procedure.—Each *S* participated in two individual sessions, with the spatial problems being administered during the first session and the anagrams during the second. The instructions for both types of problems followed a similar pattern. Instructions indicated that there would be a 7-min. time limit on each problem. It was also explained that *S* should show all his hypotheses about a problem's solution on the cards which accompanied the problem description, and that a separate card should be used for each hypothesis. This point was repetitively stressed throughout the instructions, emphasizing especially that all false starts or incorrect hypotheses should be shown. In addition, each separate problem description had attached a statement reminding *S* to show all hypotheses, one to a card.

The meaning of *hypotheses*, as used in this experiment, is determined by the operations through which they were obtained. The *S*s were told to show on the cards all ideas they might have of possible solutions to a problem, or ideas of moves toward solution. *Move* is used in the same sense as in a game like chess, i.e., as an actual step toward solving the larger problem. This meant for the spatial problems that hypotheses were ideas which were translated into drawings on cards; and, for the verbal problems, that hypotheses could be whole words or any combination of letters. The *S*s were told that they should write on the cards both letter combinations and whole words considered as possible solutions.

After the initial instructions, an example problem was administered. Before the session with spatial problems, the example was a match-stick problem of the type used by Sweeney (1953, pp. 71–72, Problem B-4), whereas the example anagram was *aehmnrs*, the name of a famous American (Sherman). As *S* worked each example problem, *E* observed his behavior, answered questions, and made sure that *S* understood how to show hypotheses. Following this, *S* was given, one at a time, the six problems for the session. He had been told that he should inform *E* when he had solved a problem. If he called an incorrect hypothesis the solution, he was told that this was not really a solution and that he should continue working.

Results

With the 7-min. time limit, the problems proved to be fairly difficult as expected. The mean number of problems solved for spatial, anagram, and both combined, respectively, was 1.95, 3.46, and 5.41. Whereas all *S*s solved at least one anagram, six failed to solve a single spatial problem.

The distribution of solutions for each of the two types of problems was divided as near the median as possible without splitting a class interval. The *S*s were also categorized into two groups on the basis of those *S*s whose anagram hypotheses were mostly whole words and those who used predominantly letter-combination hypotheses. The *S*s in the letter hypotheses group used an average of 22.65 letter hypotheses and 5.35 word hypotheses, and those in the word hypotheses group averaged 41.00 word hypotheses and 10.89 letter hypotheses, a ratio of about four to one in both cases. Two fourfold tables were constructed, one using high vs. low anagram achievement and word vs. letter hypotheses as the bases of classification, and the other employing high vs. low spatial achievement and word vs. letter hypotheses. Fisher's exact probability method was applied, and the results, as shown in Table 1, indicate for verbal problems a relation between type of hypothesis and achievement at the .008 level, and a similar trend for spatial problems at the .07 level.

Two further tests using Fisher's method were made with the spatial problems, and results for each are shown in Table 1. For the first test, high and low extremes in achievement instead of a near median split were employed, the *S*s in the extremes being determined by natural breaks near the ends of the distribution of solutions. This test yields a relation between class of hypothesis and spatial achievement which is significant at the .02 level. Interestingly,

TABLE 1

RESULTS AS FREQUENCIES IN FOURFOLD ARRANGEMENTS FOR TEST BY FISHER'S EXACT PROBABILITY METHOD

Type of Problem	Class of Hypothesis	Achievement		*P*
		High	Low	
Verbal	Word	1	8	.008
	Letter	15	8	
Spatial	Word	3	6	.07
	Letter	16	7	
Spatial Extremes	Word	2	5	.02
	Letter	8	1	
High achievement on both spatial and verbal	Word	1	6	.006
	Letter	10	2	

five of the six *S*s who failed to solve a single spatial problem use mainly word hypotheses with the anagrams. For the second test, the achievement classification was based on those *S*s who scored high both on verbal and on spatial problems vs. those scoring low on each of the two types of problems. The probability that the resulting table would occur by chance is .006.

The number of hypotheses per minute was found for each *S*, and the resulting distribution was split at the median. Fisher's test was applied to fourfold tables, with rate of hypothesis production as one variable and achievement on spatial, or verbal, or both types of problem combined as the second variable. In all three cases, there was no relation at the .10 level between achievement and number of hypotheses produced per minute.

DISCUSSION

The reason for the facilitative effect of letter hypotheses can be sought by comparing the processes involved in solving easy and difficult anagrams. The solution to anagrams depends in part upon *S*'s over-all experience with verbal material, and in part upon whether or not the letters are mixed in a way which gives readily discriminated cues, for instance, *rwfeol* vs. *erowfl*. With easy anagrams verbal cues are more readily discriminated. For difficult anagrams, where the cues are relatively abstruse, it is helpful if *S* does some educated restructuring of the letters before searching for the full word. Restructuring in terms of experience with verbal material—searching for combinations such as *un*, *less*, *ch*, or *th*, or finding larger sub-parts like the *flower* of cauliflower—tends to reduce the uncertainty associated with the anagram. In a sense, the *S* who uses letter hypotheses changes the difficult anagram into a form having the properties of an easier problem.

The positive relation between achievement and choice of letter hypotheses is in line with the assumption that one determiner of problem solving is the ability to select and use a given class of hypotheses. However, if *S*s chose to employ one or the other class of hypotheses on a strictly random basis, then those *S*s who just happened to choose the more effective class, letter hypotheses, would be expected to do better with the anagrams. This alternative possibility seems improbable in view of the results with the spatial problems. Since *S*s who used predominantly letter hypotheses also did better on spatial problems, the data provide confirmation for the ability to select and use the more appropriate class of hypotheses as a factor in problem solving.

None of the data for rate of hypothesis formation indicate a relation to problem solving. Sargent's data tentatively suggested a positive relation between achievement and rate of hypothesis production, which is not confirmed by the results of this experiment.

SUMMARY

Spatial and anagram problems were administered to explore the possibility that rate and class of hypotheses produced are

related to problem-solving achievement. Anagram hypotheses are readily separated into whole-word and letter-combination hypotheses. The latter tend to be used by *Ss* who are superior on either anagram or spatial problems, suggesting that ability to select the more effective hypothesis class is a determiner of achievement in problem solving. No relation was found between rate of hypothesis formation and problem-solving achievement.

References

HEIDBREDER, E. An experimental study of thinking. *Arch. Psychol., N. Y.*, 1924, No. 73.

SARGENT, S. S. Thinking processes at various levels of difficulty; a quantitative and qualitative study of individual differences. *Arch. Psychol., N. Y.*, 1940, No. 249.

SWEENEY, E. J. *Sex differences in problem solving.* (Contract N6onr 25125) Stanford: Dept. Psychol., Stanford Univer., 1953.

TAYLOR, D. W., & MCNEMAR, O. W. (Eds.) *Experimental studies of problem solving.* (Contract N6onr 25125) Stanford: Dept. Psychol., Stanford Univer., 1955.

(Received May 7, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

WARM-UP IN RETENTION AS A FUNCTION OF DEGREE OF VERBAL LEARNING

JUDITH E. DINNER AND CARL P. DUNCAN

Northwestern University

Irion (1949) reported that presentation of a warm-up task (color naming) just prior to recall and relearning (RL) significantly improved recall of a paired-associate verbal list. Rockway and Duncan (1952) copied Irion's method, including, among other things, an exact repetition of Irion's major conditions, but were unable to duplicate his results. None of their 11 warm-up groups performed better than the controls either on the recall or early RL trials.

Another reason for thinking that the Irion experiment may be invalid is that his one warm-up group did not show forgetting over the 24-hr. rest. As far as is known, this result, a gain rather than a loss over a 24-hr. rest in paired-associate verbal learning, has never been reported elsewhere. (At best, warm-up is expected to reduce, not eliminate, forgetting.) Thus, it seems reasonable to maintain that there is no evidence yet that warm-up facilitates recall of paired associates.

It still seemed to the writers that Irion's (1948) theoretical analysis of set and warm-up was sound. One possibility is that warm-up might facilitate recall after some degree of original learning (OL), even though not after the medium degree used by Irion, and Rockway and Duncan. Therefore, the present experiment was designed to test for effect of warm-up after different degrees of OL of paired associates.

METHOD

Materials.—The learning task was the same 15-item paired-associate list used by Irion (1949) and Rockway and Duncan (1952). Both stimuli and responses in the list are two-syllable adjectives. The list was presented on a Hull-type memory drum at a 2-sec. anticipation rate, with 4 sec. between trials. Three orders of presentation of the pairs were used.

The warm-up task was provided by eight strips of colored adhesive tape, arranged as two differently-paired 15-item lists with a 4-sec. rest between lists. Like the verbal list, a "stimulus" color appeared in the window of the drum for 2 sec., after which both the "stimulus" and its paired "response" color appeared for 2 sec. The colored strips were $\frac{3}{4}$ in. long and $\frac{1}{4}$ in. high, approximately the same size as an average adjective. The eight colors (red, green, orange, yellow, blue, brown, gray, and black) were used approximately equally often as stimuli and responses in each list, and were paired so that no two were in the same stimulus-response relationship more than once in either list. The lists of colors were mounted on the right side of the same tape as the verbal list. The change from colors to adjectives was made in the 4-sec. intertrial interval.

Conditions.—Three degrees of OL, 3, 8, and 15 (one errorless trial) correct anticipations, were used. An experimental (warm-up) and a control (no warm-up) group learned to each degree on Day 1 and returned 24 hr. later for either warm-up and relearning for 10 trials (experimental) or merely relearning for 10 trials (control). The medium degree of OL (8 correct) is close to the average mastery achieved in the Irion, and Rockway-Duncan studies.

Subjects.—The *S*s were 96 male and female students from introductory psychology classes at Northwestern University, rewarded with one examination point for each day served. They were assigned in turn to six groups of 16 *S*s each. Replacements were obtained for nine *S*s who were dropped for failure to make a single correct response by the end of the fifth OL trial. No *S* had to be dropped for exceeding the maximum criterion: errorless performance on the fifth OL trial. No *S* was dropped for deliberate rehearsal during the 24-hr. retention interval; all denied such rehearsal.

TABLE 1

MEAN TRIALS TO REACH VARIOUS CRITERIA OF CORRECT RESPONSES IN ORIGINAL LEARNING

Criterion	Experimental		Control	
	Mean	σ_M	Mean	σ_M
3/15	5.88	.52	5.88	.76
8/15	12.94	1.50	13.69	1.77
15/15	24.50	2.70	24.56	2.82

Procedure.—On Day 1, *S* was read instructions for paired-associate learning, learned the list to the appropriate criterion, was cautioned against rehearsal, and dismissed.

On Day 2, *S*s in the three experimental groups were read instructions for color naming. These instructions emphasized that *S* was not to learn the colors, but was to wait until the shutter raised and then say aloud the name of the second color of a pair. Saying aloud the name of the second color was emphasized. Two trials of color naming were given, one on each of the two differently paired lists of colors. Immediately after completing the second warm-up trial *S* was instructed that the words he studied the day before would be presented for 10 trials, and that he was to make as many correct responses as he could the very first time through the list. The time between completing the warm-up task and the beginning of recall was less than 10 sec.

The control groups were given the same instructions concerning recall and RL of the verbal list as the experimental groups, but were given no warm-up.

RESULTS

Original learning.—The mean number of trials for each group to reach the various OL criteria (3, 8, or 15 correct anticipations) is shown in Table 1. In terms of trials to reach the criterion, the three pairs of groups are widely separated. Within each pair the warm-up and control groups

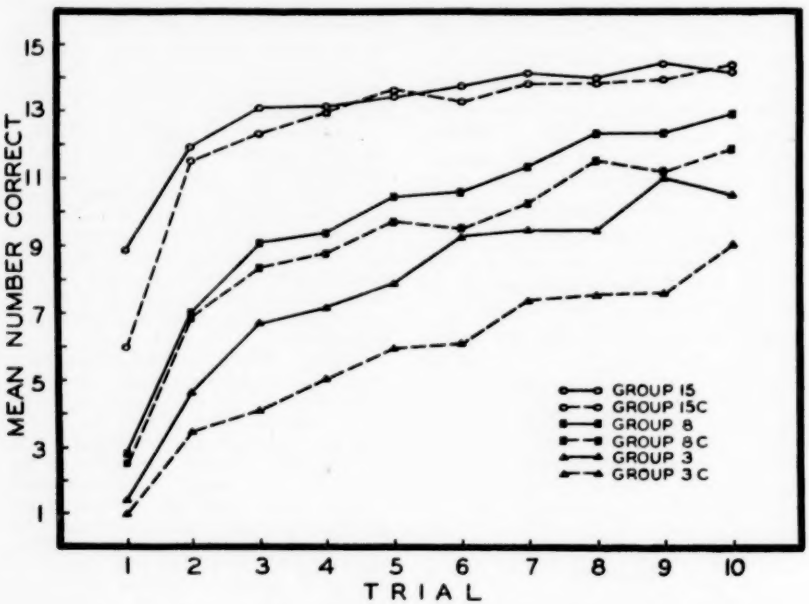

FIG. 1. Mean number correct on each relearning trial for each group.

TABLE 2

ANALYSIS OF NUMBER CORRECT ON RECALL

Source	*df*	*MS*	*F*
Warm-up (W)	1	33.85	1.97[a]
Learning (L)	2	338.10	19.70[a]*
W × L	2	17.16	4.70*
Within groups	90	3.65	

[a] Using interaction as error.
* $P = .05$.

are fairly well matched; the largest difference is less than one trial. The mean number of words correctly anticipated on the last OL trial was, of course, exactly 15 in the high OL groups. For warm-up and control *Ss* respectively, the corresponding means were 8.88 and 8.69 in the medium OL groups, 3.62 and 3.56 in the low OL groups.

Recall.—Performance of each group on each RL trial is shown in Fig. 1. By inspection, it appears that warm-up improves recall (first-trial performance) mainly for the high degree OL.

An analysis of variance on the recall scores (warm-up vs. control as one variable, degree of OL as the other) is summarized in Table 2. Since the interaction between warm-up and OL was significant, it was used as the error term for the main effects. (Using within groups as error, $F = 9.27$ for warm-up, 92.63 for OL, both significant at the 1% level.) As Table 2 shows, OL was significant, using interaction as error, but warm-up was not. Because of the significant interaction, each experimental group was compared to its corresponding control group by t. Between Groups 3 and 3C (C indicates control), $t = .90$; between Groups 8 and 8C, $t = .42$, but between Groups 15 and 15C, $t = 3.27$, which for 30 *df* is significant beyond the 1% level. Thus, as Fig. 1 suggests, recall was improved by warm-up only after the high degree of OL. The improvement was fairly large: the mean words recalled by the high OL *Ss* was 9.0 in the warm-up group, 6.0 in the control group.

Relearning.—Figure 1 suggests that, in direct contrast to the results on recall, warm-up improved performance in RL beyond the first trial mainly in the low, and to some extent in the medium, degrees of OL. An analysis of variance on total score over all 10 trials (i.e., including the recall trial) is summarized in Table 3. In this case the interaction term was not significant, but warm-up, and, of course, OL, were significant at the 1% level. Using the within-groups term from the analysis of variance to compute an error term, the t between Groups 3 and 3C was 3.79, which is significant beyond the 1% level. Between Groups 8 and 8C, t was 1.4, between Groups 15 and 15C, t was 1.00, neither of which is significant.

Because warm-up seemed to have a different effect on recall than on RL, another analysis of variance was done on total score over the last nine trials of RL, omitting the recall trial. The results were much the same as those in Table 3. The F was 10.84 for warm-up, 109.63 for OL, both significant at the 1% level. The F for interaction (2.97) was not significant. Also, t tests, corresponding

TABLE 3

ANALYSIS OF NUMBER CORRECT OVER ALL 10 RELEARNING TRIALS

Source	*df*	*MS*	*F*
Warm-up (W)	1	2981.51	12.76**
Learning (L)	2	29,928.01	128.06**
W × L	2	528.14	2.26
Within groups	90	233.70	

** $P = .01$.

to those based on Table 3, yielded the same results; the t values were 3.82 ($P < .01$), 1.41, and .48 for the low, medium, and high degrees of OL, respectively.

Discussion

For 24-hr. retention of paired adjectives, two facilitating effects of warm-up were found, one with high OL, a different effect with low OL. With medium OL, warm-up had no effect.

The fact that warm-up did not significantly facilitate either recall or RL after a medium degree of OL confirms the findings of Rockway and Duncan (1952), and again contradicts the Irion 1949 study. It seems likely that Irion's results (facilitation of recall by warm-up after a medium degree of OL) represent a sampling error and should be discounted. It is also worth noting that the procedure for achieving medium OL in the present study (training to a criterion) versus that used in the Rockway-Duncan study (constant number of trials) led to much greater forgetting in the present study (mean loss of 6.10 words versus a mean loss of 1.70 words for Rockway and Duncan), yet warm-up had no effect on recall after medium OL in either study.

Warm-up significantly facilitated recall, but not later RL, after the high degree of OL. This finding seems in accord with Irion's (1948) theoretical account of what warm-up should do. Briefly, over a forgetting interval there is loss of set; a warm-up activity, given just prior to recall, should reinstate the set and thereby improve recall. From the present results it appears that this "recall effect" of warm-up is limited to high degrees of OL. The effect also does not persist throughout RL, probably because the recall trial warms up the control group (see Fig. 1).

Warm-up facilitated RL as a whole, but not recall, after the low degree of OL. In interpreting this it should be noted that in the low OL groups the 10 RL trials should really be called OL rather than RL (see Fig. 1). On Day 2 only one trial was needed to return these *Ss* to their original level (3 correct); remaining trials were devoted to acquiring, not re-acquiring, habit strength. This suggested examination of studies in which warm-up facilitated OL of verbal paired associates. One of these studies, Thune (1950), presented trial-by-trial learning curves. Comparison of the curves in the Thune study (1950, Fig. 2) with the bottom two curves of the present Fig. 1 shows that in both studies the warm-up group was slightly superior to the control group on the first trial given after presentation of the warm-up task. In each case this superiority quickly increased, then remained fairly constant until practice was terminated. Thus, the facilitating effect of warm-up was greater in later than in earlier trials in both studies. (Thune concluded that the facilitation was greater in earlier than in later trials because he compared slopes of, rather than vertical distances between, curves.)

The effect of warm-up found in OL by Thune, and in RL after low OL in the present study, might be interpreted as follows. Even though the experimental group has developed the proper set, via the warm-up task, there is apparently no way in which the set could immediately produce a large facilitating effect. Habit strength cannot be raised from a low to a high level in one trial by set alone. At the same time, the initial slight facilitating effect of set in the experimental group can easily persist over several trials because the control group is only gradually acquiring set. Furthermore, if the effect of set is to increase, even slightly, the rate of gain, the superiority of the warmed-up group would not only persist, but actually increase in amount, at least for a time.

Finally, it should be noted that the two quite different effects of warm-up that were found, a "rate-of-gain effect" with low OL and a "recall effect" with high OL, probably resulted from the use of only three widely separated degrees of OL. Since OL is a continuous

variable, the effect of warm-up should also vary continuously as a function of degree of OL; the present study may have identified the extremes of a dimension of warm-up effects. If such a dimension exists, it would help explain why warm-up had no effect after the medium degree of OL. At this level of mastery, the rate-of-gain effect of warm-up, characteristic of low OL, has begun to disappear (see Fig. 1) before the recall effect, found with high OL, starts to occur.

Summary

The study attempted to resolve contradictory findings concerning the effect of warm-up on recall of verbal paired associates. Six groups of *Ss* were first given practice on a list of 15 paired adjectives. Degree of original learning was varied by practicing one pair of groups to a criterion of three correct responses, a second pair to eight correct, and a third pair to one errorless trial. Just prior to recall and relearning, 24 hr. later, one group in each pair was given two trials on a warm-up task, consisting of color naming.

The results were:

1. Warm-up significantly facilitated recall only after the highest degree of original learning; it had no effect on recall after the medium or low degrees of learning. For this and other reasons it was suggested that a previous report of facilitation of paired-associate recall by warm-up after a medium degree of learning should be rejected.

2. Warm-up significantly facilitated overall relearning after the low degree, but not after the medium or high degrees, of original learning. In this case the facilitating effect of warm-up appeared as an increased rate of gain of habit strength in a manner similar to that which has been found when warm-up is given prior to original paired-associate learning.

It was suggested that since only widely separated and arbitrary categories of OL were used, the two quite different effects of warm-up that were found probably represent the extremes of a dimension of warm-up effects.

References

Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55,** 336–341.

Irion, A. L. Retention and warming-up effects in paired-associate learning. *J. exp. Psychol.*, 1949, **39,** 669–675.

Rockway, M. R., & Duncan, C. P. Pre-recall warming-up in verbal retention. *J. exp. Psychol.*, 1952, **43,** 305–312.

Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40,** 423–438.

(Received May 8, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

MEANINGFULNESS AND ARTICULATION OF STIMULUS AND RESPONSE IN PAIRED-ASSOCIATE LEARNING AND STIMULUS RECALL [1]

RAYMOND G. HUNT [2]

University of Buffalo

It seems fairly well agreed that the general effect of meaningfulness is to facilitate the acquisition of verbal materials. However, the question of the relative effects of meaningfulness of the stimulus and response members in paired-associate learning remains an open one. The empirical literature bearing upon it is by no means clear. Noble and McNeely (1957) Cieutat, Stockwell, and Noble (1958) have found both stimulus and response meaningfulness to facilitate learning but conclude that response factors are of greater importance. Kimble and Dufort (1955) report similar data but arrive at the opposite conclusion. Sheffield (1946) found response meaningfulness, but not stimulus meaningfulness to facilitate learning. Mandler and Campbell (1957) found consistent and clear-cut effects only in regard to response meaningfulness or association value. Further investigation in this area is clearly in order.

Linked to the question of the roles of stimulus and response factors in paired-associate learning is the question of what psychological mechanisms may be responsible for the effects observed. Noble has suggested that variations in the motor skill of pronunciation may be one of the gross variables underlying the relative ease of learning meaningful verbal material. Such a notion is consistent with Noble's emphasis upon the primary role of response factors in paired-associate learning and suggests the hypothesis that practice in pronouncing the stimulus words during the learning of paired-associates might facilitate their acquisition. Such facilitation could occur by virtue of the pronunciation process providing more stable internal cues which then mediate the acquisition of "associations."

Finally, Bugelski (1957), for one, in an effort to further specify the role of stimulus members of paired associates, has proposed that, in the paired-associate procedure, the stimulus term plays only a minimal role. From his theoretical analysis Bugelski suggests that its usual function is simply to provide a "cue" occasioning the performance of the response term. Given such a state of affairs, any discriminable portion of the stimulus term can suffice. In other words, *S* need not learn the term completely nor even pay much attention to it as a word. As a simple and preliminary test of this hypothesis Bugelski proposes a "backward" recall procedure like that used by Feldman and Underwood (1957) to demonstrate the acquisition of R-S associations in the course of learning S-R associations. It would be expected that *S* will be unable to recall stimuli

[1] This paper is based upon a dissertation submitted to the Department of Psychology at the University of Buffalo in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to thank B. R. Bugelski, chairman, and the other members of his doctoral committee for their advice and assistance. Thanks are also due K. H. Kurtz for his help in preparing this paper.

[2] Now at Washington University, St. Louis, Missouri.

of low meaningfulness since they are initially unfamiliar and are not learned as units during the acquisition trials. On the other hand, meaningful stimuli should be recalled readily as these have already been learned prior to the experiment and so will tend to be associated as wholes with the response members.

METHOD

Experimental design.—Three main experimental variables formed the basic structure of the experiment: (*a*) stimulus meaningfulness (high or low); (*b*) response meaningfulness (high or low); and (*c*) articulation or nonarticulation of the stimulus word (response words were necessarily articulated in all conditions).

Four main experimental conditions were employed as follows: High-High (high stimulus meaning—high response meaning; Low-Low (low stimulus meaning—low response meaning); High-Low (high stimulus meaning—low response meaning); Low-High (low stimulus meaning—high response meaning).

Each of these conditions was divided into two subconditions, in one of which *Ss* overtly verbalized the stimulus words during learning and in the other of which they did not. While it is likely that some *Ss* in the nonarticulation conditions verbalized the stimulus subvocally, it is unlikely that anything approaching the 100% vocalization of the articulation groups obtained. It is questionable in any case whether subvocal articulation is equivalent to overt vocalization.

Apparatus.—A device adapted to the presentation of verbal materials for paired-associate learning and previously described by Bugelski and Scharlock (1952) was used. This apparatus consists essentially of a screen with an aperture behind which revolves a sectioned disk. The revolving disk exposes words at predetermined intervals. All words were block printed in India ink on 5 × 8-in. white cards with letters 4 mm. high by 3 mm. wide. The cards bearing the paired words were stacked behind the aperture and were changed manually. Individual S-R pairs were withdrawn when the criterion of two successive correct anticipations was attained, thereby controlling overlearning.

The stimulus word of each pair was first exposed alone for 2 sec., then both words of the pair were exposed for 2 sec.; the interval between successive pairs was 4 sec.

Material.—Eight paired-associate lists were used, two in each of the four main experimental conditions. Each consisted of 10 pairs drawn from Noble's (1952) list of disyllabic nouns and paralogs scaled for meaning (*m*). A High-High list was constructed by selecting 20 words at random from the 34 in Noble's list having *m*-scale values greater than 5.50. In one such list the odd-numbered words formed the stimulus units and the even-numbered words the response units. A second High-High list was formed by reversing these positions.

Two Low-Low lists were derived in identical fashion from the 25 words in Noble's list having *m*-scale values less than 2.00.

Two High-Low and two Low-High lists were constructed by appropriate combinations of the words in the High-High and Low-Low lists.

The usual precautions were taken to avoid possible biasing effects peculiar to individual pairs.

Finally, a six-pair practice list was constructed. This consisted of geometric figures as stimuli and colors as responses. It was assumed that such a list would effectively acquaint *S* with the procedure and apparatus. Each *S* received a minimum of five practice trials.

Subjects.—Eighty volunteer undergraduate students (54 men, 26 women) enrolled in summer session courses at the University of Buffalo served as *Ss*. They ranged in age from 17–30 yr. (mean 21, *SD* 3 yr.). Approximately one-fourth of these *Ss* had participated previously in verbal-learning experiments.

Procedure.—Twenty *Ss* were assigned to each of the four main experimental conditions. Within each condition 10 *Ss* vocalized both S and R terms while the remaining 10 *Ss* vocalized only the R terms. Five *Ss* in each of these subconditions learned one of the two experimental lists for that condition. All *Ss* were assigned randomly to the experimental conditions and no *S* served in more than one condition.

Learning was by the anticipation method with correction on all but the first run through the list, during which *S* simply pronounced the appropriate terms as they appeared. Common pronunciation was agreed upon at this time. The technique of correction varied according to *S*'s treatment condition. The *Ss* in the articulation conditions corrected by pronouncing both the S and R words whereas *Ss* in the nonarticulation conditions corrected by pronouncing only

the correct R word. The *Ss* were instructed to avoid wild guessing.

One run through a list of 10 pairs constituted a trial, the cards being shuffled after each trial to control serial effects. The intertrial interval was about 15 sec.

Learning was measured in terms of median trial to criterion for each *S*. This score was obtained by determining the trial on which each of the 10 pairs making up the experimental list was learned, placing these in rank order and then determining their median.

Immediately following the learning trials each *S* was tested for recall of the *stimulus* words of the pairs just learned. The response words of these pairs were presented for 4 sec. each. The *Ss* in the articulation groups pronounced each word as it was shown and then attempted to reproduce, either by pronouncing or by spelling, the word which had been paired with it. The *Ss* in the nonarticulation groups simply viewed the word without pronouncing it. Recall was measured in terms of number of errors, an error being defined as anything less than perfect reproduction. Each *S* was encouraged to report whatever he could recall of the stimulus term in the event he was unable to reproduce it completely.

Results

Acquisition.—The mean acquisition scores for the four main experimental groups were 3.19, 5.10, 7.18, and 8.31 for the High-High, Low-High, High-Low and Low-Low groups, respectively. These mean scores are consistent with the hierarchy of effects predicted by Noble and McNeely (1957) and reported by (Cieutat, Stockwell, and Noble (1958).

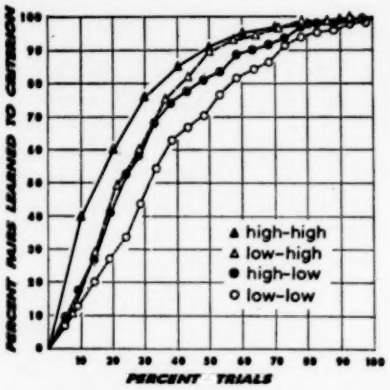

FIG. 1. Cumulative acquisition curves for the four main experimental conditions.

A similar conclusion might be drawn from the learning curves presented in Fig. 1. Because of the procedure used in the learning, not

TABLE 1

ANALYSES OF VARIANCE OF STIMULUS, RESPONSE, AND ARTICULATION FACTORS IN PAIRED-ASSOCIATE ACQUISITION

Analysis	Source[a]	*df*	Acquisition		Recall	
			MS	*F*	*MS*	*F*
Response	Treatment	1	265.72	141.34**	40.61	16.78*
	Error (Lists within treatments)	2	1.88		2.42	
Stimulus	Treatment	1	25.66	1.85	374.11	239.81**
	Error (Lists within treatments)	2	13.88		1.56	
Articulation[a]	Treatment	1	0.00	—	.12	—
	Error (Lists × Articulation)	4	82.21		13.66	

* $.05 < P < .10$.
** $P < .01$.
[a] Only the sources of interest for present purposes are selected from the complete analysis for presentation.

all *S*s received the same number of trials. Consequently the curves in Fig. 1 are "Vincent-type" adjusted for unequal trials by a method described by Woodworth and Schlosberg (1954). It can be seen that each of these curves conforms to the usual negatively accelerated growth curve and that the respective steepness of slope of the four curves is consistent with the Noble-McNeely hypothesis.

However, as reported in Table 1, analyses of variance of the effects of stimulus and response meaningfulness reveal that only variations in response term meaningfulness produced a statistically significant effect upon acquisition rates. Neither stimulus meaningfulness nor articulation had any significant effect in the present experiment. The data reported in Table 1 were obtained by performing independent analyses of the effects of the three main experimental variables.

In the case of the stimulus and response analyses the error term used to test experimental effects is based upon sampling variation among lists of similar *m* value. Since the same lists were used under articulation and nonarticulation conditions, the list-by-articulation interaction was employed as an error term for evaluation of the effects of articulation. Inasmuch as *S* variability was confounded with list variability, each of these analyses permits generalization both to a population of *S*s and to a population of lists.

Recall.—The mean numbers of errors in stimulus recall were 1.80, 5.70, 2.80, and 4.30 for the High-High, Low-Low, High-Low and Low-High groups, respectively. Results of analyses of variance of errors in stimulus recall are presented in Table 1. As predicted by Bugelski, variations in stimulus meaningfulness produced a statistically significant effect upon stimulus recall. Similar variations in response meaningfulness produced an effect which approaches significance. Stimulus articulation, however, had no significant effect upon stimulus recall.

Discussion

The main results of the present study are consistent with those presented by Sheffield (1946). Differential acquisition rates appear to be associated primarily with response dimensions of the paired-associate process. It seems quite possible that the diverse findings reported in the literature concerning the effects upon acquisition of varying stimulus meaningfulness may be a function of variations from experiment to experiment in sampling from the range of meaningfulness, or association values. Preliminary analysis (not reported) of the present data revealed as untenable any assumption of an absence of variance in the data attributable to differences among stimulus materials of similar *m* value. It is essential that this variance be accounted for in the error term used to test experimental effects. The positive effects upon acquisition sometimes obtained by increasing the meaningfulness of the stimulus member of the paired associate may reflect sampling variations attributable to materials rather than true effects. Further experimentation in this area should probably involve careful sampling from a wide range of *m*, or association, values, while at the same time closely controlling variations on the response side of the paired associate. In any case, judging from the results presented here, in conjunction with those of Sheffield, and in view of the inconsistent effects reported by Mandler and Campbell (1957), it seems likely that variations in the meaningfulness of the stimulus member are not apt to produce consistent effects upon paired-associate acquisition.

The fact that recall is better for meaningful stimuli is consistent with Bugel-

ski's hypothesis of the "minimal cue" function of the stimulus member in such learning. Moreover, the fact that increasing response term meaningfulness appears to produce greater facilitation of paired-associate learning than does increasing stimulus meaningfulness can also be accommodated within the Bugelski hypothesis. Such a finding would be expected since it is postulated that only the response member of the pair need be performed in its entirety. Therefore, meaningfulness of that unit would be expected to aid learning more than would comparable degrees of stimulus meaningfulness.

Some further support for the Bugelski hypothesis can be taken from the frequency of partial recall in the high- and low-stimulus-meaning conditions. Partial recall, defined as some part of the appropriate term but less than the whole, accounted for 37% of the errors in recall for the low-stimulus-meaning condition but only 8% in the high-stimulus-meaning condition. These differences were primarily a result of a decrease in the percentage of intrusion errors from 49% in the high-stimulus-meaning condition to 17% in the low-stimulus-meaning condition (the percentage of failures was 46% in both). These findings would be expected from Bugelski's thesis which assumes *Ss* in the low meaning condition to be discriminating as a cue only a fragment from the general topography of the total stimulus. The *S* would thus be expected to be prepared to perform only that fragment rather than the whole word.

The present failure to demonstrate any facilitation of learning as a function of stimulus articulation does not necessarily contradict an emphasis upon motor factors as mediating the facilitative functions of meaningful material. It may simply be that *S* discriminates fragments of the stimulus word with which he is "familiar"; a process which could operate along linguistic lines. The larger process of "integrating" an instrumental response to the stimulus word could be a process concomitant with but largely independent of the formation of an association between some fragment of that word and some other unit (the response word).

Finally, the present data bear upon the "backward" recall findings of Feldman and Underwood (1957). These investigators reported the correct recall of stimulus items following paired-associate learning to be of the order of 50%. Nonsense syllables were used as stimuli and adjectives as responses. Their experimental lists, therefore, were comparable with the Low-High condition in the present experiment. The percentages of stimulus items correctly recalled (perfect reproduction) in the present study were 99, 80, 54, and 32 for the High-High, High-Low, Low-High and Low-Low conditions, respectively. The percentages of stimuli recalled by comparable groups in the two studies are almost identical, indicating the generalization that stimulus recall will be approximately 50% following the learning of similarly constituted paired associates. As the stimulus or response terms vary in meaningfulness, a concomitant increase or decrease, respectively, in recall may be expected.

Summary

Eighty *Ss* learned lists of paired associates in which the stimulus and response terms varied in meaningfulness (*m*). Four main experimental conditions were employed (High-High, High-Low, Low-High, and Low-Low). In addition half of the *Ss* in each condition articulated the stimulus word during learning while the other half did not.

Each *S* learned by the anticipation method with correction to a criterion of two successive errorless anticipations. Immediately after the acquisition trials *Ss* were given a "backward" recall test.

The main results were:

1. Variations in response meaningfulness produced a significant effect upon acquisition. Stimulus meaningfulness and articulation produced no significant effects.

2. Stimulus meaningfulness produced significantly greater *stimulus* recall, and the effect of response meaningfulness on stimulus recall approached significance. Stimulus ar-

ticulation had no significant effect upon stimulus recall.

3. Learning curves for each of the four main experimental conditions were negatively accelerated and conformed to the hierarchy of effects predicted by Noble and McNeely.

It is concluded that the results of the present experiment generally support an S-R motor patterning view of paired-associate learning.

References

Bugelski, B. R. The role of familiarity in the transfer of training situation. Paper read at Amer. Psychol. Ass., New York, Sept., 1957.

Bugelski, B. R., & Scharlock, D. F. An experimental demonstration of unconscious mediated generalization. *J. exp. Psychol.*, 1952, **44,** 334–338.

Cieutat, V. J., Stockwell, F. E., & Noble, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m, m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56,** 193–202.

Feldman, S. M., & Underwood, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53,** 11–15.

Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50,** 361–368.

Mandler, G., & Campbell, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54,** 269–273.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59,** 421–430.

Noble, C. E., & McNeely, D. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53,** 16–22.

Sheffield, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale Univer., 1946.

Woodworth, R. S., & Schlosberg. H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

(Received May 8, 1958)

Journal of Experimental Psychology
Vol. 57, No. 4, 1959

STIMULUS GENERALIZATION AFTER TRAINING ON THREE STIMULI: A TEST OF THE SUMMATION HYPOTHESIS [1]

HARRY I. KALISH AND NORMAN GUTTMAN

Adelphi College and Duke University

A previous study in this series which explored the problem of combination of stimulus generalization gradients after equal training on two stimuli (Kalish & Guttman, 1957) did not conclusively differentiate between the Hullian summation hypothesis and the hypothesis that the total generalization gradient is the noninteractive resultant of two independent gradients (nonsummation hypothesis). Both hypotheses provided fairly reasonable accounts of the empirical data. Moreover, the analysis suggested that a more powerful test of the summation hypothesis could be achieved by training to three test stimuli arranged close enough to ensure that the center stimulus received a considerable degree of excitation from the gradients of the adjacent stimuli. Under these conditions the predictions from Hullian theory would be unequivocal. Specifically, training to three spectral stimuli spaced 10 mμ apart should generate an over-all gradient with pronounced rise in rate of responding to the center stimuli. In this case, the alternative predictions should be sufficiently different to distinguish between the two hypotheses.

METHOD

Subjects.—The *S*s were 10 experimentally naive White Carneau pigeons.

Apparatus.—The apparatus, a modified version of the Skinner automatic key-pecking apparatus, was the same as that used in previous investigations by the authors (Guttman & Kalish, 1956, Kalish & Guttman, 1957).

Procedure.—The training and testing procedure was essentially the same as that previously employed (Kalish & Guttman, 1957), except that each *S* was trained to peck at three monochromatic stimuli: 530, 540, and 550 mμ. The *S*s were maintained at 80% of normal body weight determined under ad libitum feeding.

Initial training was under continuous reinforcement with food. A total of 120 reinforcements (40 for each stimulus) were given during two training sessions, and the stimuli were changed every five reinforcements. Following continuous reinforcement, 10 daily 30-min. periods of variable interval reinforcement were given. Each daily session consisted of two 5-min. presentations of of each of the three stimuli. Each session was divided into 1-min. intervals by 10-sec. blackouts. The change of stimuli was accomplished during the blackouts.

During the generalization test, each *S* was presented with 11 nonreinforced test stimuli consisting of the following series: 500, 510, 520, 530, 535, 540, 545, 550, 560, 570, and 580 mμ. These test stimuli were randomized within a series, and 12 different random series were presented to each *S*. Each stimulus presentation was 30 sec., and was followed by a 10-sec. blackout. On the day following this test, a second generalization test identical to the first was administered.

RESULTS

The mean generalization gradients obtained in the first and second tests are shown in Fig. 1. Beyond the region of the conditioned stimuli, both curves show the decrement in response strength which is characteristic of the generalization phe-

[1] This study was supported by grants M-1459 and MH-1002 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Milton Kornrich for his assistance in this research.

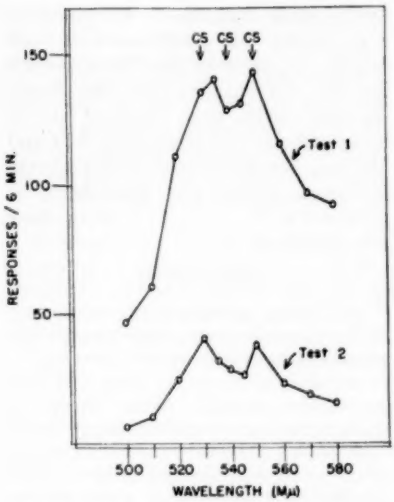

FIG. 1. Mean generalization gradients for Tests 1 and 2.

nomenon. Within the region of the conditioned stimuli, however, unanticipated differences appear. Specifically, the rate of responding to the middle CS (540 mμ) is below the rate to the other stimuli which is contrary to expectations from the summation hypothesis. In order to obtain a precise appraisal of alternative theories, comparisons must be made between theoretical functions derived from both the summation and nonsummation hypotheses and the empirical curve. As in the previous treatment (Kalish & Guttman, 1957) the question of theoretical interpretation depends upon whether certain assumptions permit the derivation of an appropriate hypothetical function.

The theoretical curve representing the Hullian summation hypothesis in Fig. 2 was derived using the same assumptions and procedure employed in the analysis of the two-stimulus problem. Specifically, three equal and smaller single stimulus gradients were taken from the previously determined "generalization surface" (Guttman & Kalish, 1957, Fig. 4) and were combined according to the Hullian rule of exponential addition (Hull, 1943). The values of the hypothetical summated curve are governed principally by one parameter, the height of the underlying gradients at their respective CS's. This parameter was determined by generating the set of summated curves resulting from various height parameters within a previously estimated range, and by computing for each summated hypothetical curve the sum of the differences between the curve and the empirical data. A plot of the sum of the differences against the height parameter goes through a single definite minimum, and the best-fitting hypothetical curve shown in Fig. 2 corresponds to this minimum.

A statistical test of the goodness of fit of the hypothetical to the observed data by means of analysis of variance (Lindquist, 1953, pp. 344–347) was

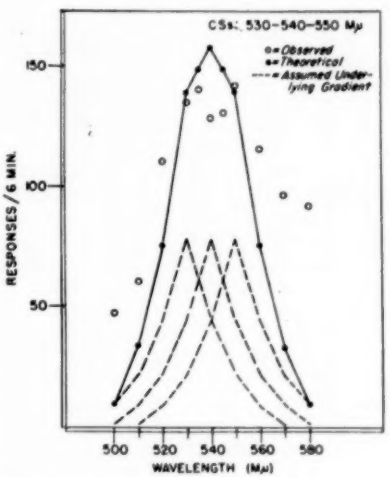

FIG. 2. Summation hypothesis: theoretical curves and observed values.

applied to the curve after excluding the values 570 and 580 mμ. These values were excluded from the analysis because their conspicuous divergence at noncritical points would tend to bias the analysis of the center points toward rejection of the goodness-of-fit hypothesis. This would place a heavier burden on the summation hypothesis than is actually warranted since, in this instance, the hypothesis emphasizes changes around the center of the gradient near the training values. The results of the analysis indicated a significant departure from hypothesis ($F = 6.59$; 8, 72 df; $P < .001$). Further examination of the individual points revealed that the departure from hypothesis was due to the values 500, 520, 540, and 560 mμ, all of which indicated differences between the observed and theoretical means significant at approximately the .05 level.

The theoretical curve shown in Fig. 3 is simply the outline of three independent, noninteracting gradients taken from the generalization surface at the same heights as the empirical CS values. A goodness-of-fit test was also applied to these data excluding the CS values, since the hypothetical curve was required to fit these points, and excluding 570 and 580 mμ, for the reasons given above. In this instance, the hypothesis of goodness of fit was not rejected, since $F < 1.00$.

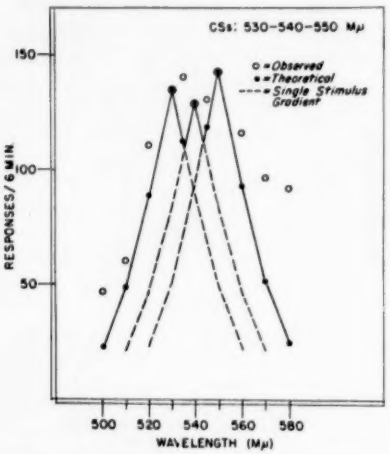

FIG. 3. Nonsummation hypothesis: theoretical curves and observed values.

DISCUSSION

The results of the goodness-of-fit tests of the two hypotheses indicate that the nonsummation hypothesis provides a better description of the empirical generalization gradient. With respect to the summation hypothesis there is no evidence for a rise at the center stimulus which would be predicted by the theory. The failure to obtain the predicted rise resulting from a summation process might be attributed to the flattening of the gradient in the center due to the accumulation of inhibition during the nonreinforced test trials. This presupposes that a greater amount of inhibition would accrue to the center stimuli because of a summation process and further suggests that the stimuli in the center would decline at a faster rate relative to the adjacent stimuli. An analysis of the gradients at successive stages of testing, however, fails to support such a possibility. Four separate gradients were obtained by separating the generalization test data into four blocks of three test series each. A test of trend yielded an $F < 1.00$ for the Test-series and Stimuli (530, 535, 540, 545 and 550 mμ) interaction which indicates that the responses to all stimuli were being reduced at approximately the same rate.

Not only does the summation theory fail to be confirmed by the absence of a rise at the center stimulus, but the curve developed on the basis of this theory becomes systematically too narrow on the wings with the inclusion of further underlying gradients within a circumscribed stimulus range. As a consequence of adding more stimuli, the

available response strength in the center of the curve must be divided among a larger number of progressively smaller hypothetical gradients. This entails a reduction in the number of responses in the wings of the summated curve, and as the previous detailed analysis has shown, the empirical curve is too broad for the best-fitting summated curve.

It should be borne in mind that the numerical differences between the two hypotheses under consideration are extremely small as may be seen in Fig. 4. The experimental differentiation between these two views is technically very difficult, since for a broad range of situations, it is possible to construct a summation hypothesis which very closely approximates the nonsummation hypothesis. This point is illustrated by the curves in Fig. 5 to represent alternative hypotheses for the situation of three CS's spaced 5 mμ apart. This reduction in interstimulus distance results in a very small increase in the expected rise for the center stimulus, while now the differences in the wings of the gradient become almost negligible. An experiment which attempted to maximize differences through the use of this design in order to obtain a more exact test of the summation hypothesis would gain very little in power. In fact, assuming the means and variances of the present experiment and equality of the empirical peaks, the difference at the center value would not be significant at the .05 level. Thus it would seem that the maximal experimental difference between the hypotheses can be achieved by increasing the number of CS's rather than by reducing their spacing on the stimulus continuum.

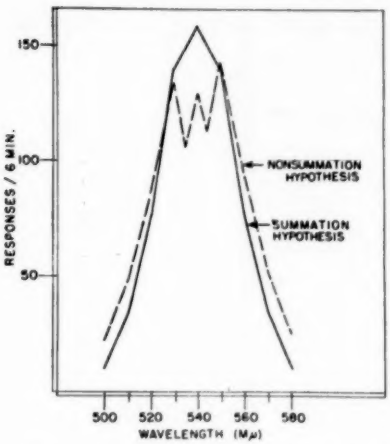

FIG. 4. Comparison of the summation and nonsummation hypothesis for the three-stimulus problem with a spacing of 10 mμ among the CS values.

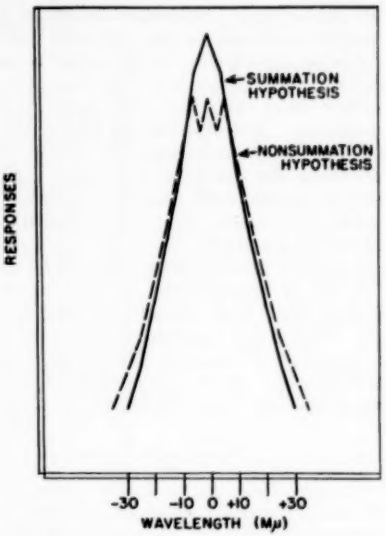

FIG. 5. Comparison of the summation and nonsummation hypothesis for the three-stimulus problem with a spacing of 5 mμ among the CS values.

Although it is not possible to reject the nonsummation hypothesis as a description of the mean empirical gradient, it is necessary to maintain certain reservations about the adequacy of this hypothesis as an account of the individual case. The flatness in the center region of the mean gradient appears to be characteristic of relatively few individual *S*s. Typically, the animal shows a decided bias in favor of one or two

of the equally reinforced stimuli. A parallel is found in the previous two-stimulus experiment, where equal rates to the two stimuli were the exception rather than the rule for individual gradients. The flatness in the center of the mean three-stimulus curve is therefore the result of averaging individual curves with maxima at several values. A more complete description of these situations may have to take account of the possibility that the *S* is forming a spontaneous discrimination among the three stimuli (or between the two, when two are used). Such a pattern of behavior may well be facilitated by the intermittent reinforcement schedule employed in these studies.

Summary

An experiment was performed to test the Hullian hypothesis of summation of generalization gradients by training a group of pigeons ($N = 10$) to peck at three monochromatic stimuli spaced 10 mμ apart. Goodness-of-fit tests to the theoretical curves disclosed that the empirical data were not well described by the summation hypothesis. The non-summation hypothesis provided a more adequate description of the obtained mean generalization gradient.

Implications of the Hullian model for this case were discussed, and disparities between theoretical and individual gradients were noted.

References

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received for early publication November 6, 1958)

OPUS #1

An event of importance for

NEW AUTHORS ONLY

OPUS #1 is the title of a new Philosophical Library 1959 book series which will publish the deserving manuscripts of new authors only.

OPUS #1 has two basic requirements: The author must have something to say and know how to say it.

OPUS #1 will be dedicated to the humanities and will include works in the fields of Art, Literature, Philosophy, Religion, Sociology, History, Psychology and Psychiatry.

OPUS #1 books will be printed on special, fine antique paper, and artistically cloth-bound for library use.

Only complete book-length manuscripts accepted.

Inquire for full details

Philosophical Library, Publishers

Dept. OP-111

15 East 40 St., N.Y. 16, N.Y.

Schizophrenia Manfred Sakel	5.00
Epilepsy Manfred Sakel	5.00
The Analysis of Dreams Medard Boss	6.00
Group Psychoanalysis B. Bohdan Wassell	3.75
The Neuroses and their Treatment Edward Podolsky	10.00
Psychotherapy and Society Wladimir Eliasberg	6.00
The New Chemotherapy in Mental Illness H. L. Gordon	12.00
Principles of Self-Damage Edmund Bergler	6.00
The New Psychiatry Nathan Masor	3.75
Experimental Psychology Ivan Pavlov	7.50

Philosophical Library, Publishers

15 East 40 Street, New York 16

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Journal of the German Psychological Association

Editors for Germany

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Editor for Austria	Editor for Switzerland	Redaction
PROF. DR. H. ROHRACHER, Vienna	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Vol. 5, No. 3

ARTHUR L. BENTON. **The Benton Test**

WOLFGANG VON MOERS-MESSMER. **A Contribution on the Psychology of Form Gestalt**

EBERHARD ULICH. **Question of the Five-Day-Week in the Schools**

HORST MYERHOFF. **Experimental Research on the Psychology of Tubercular Diabetics**

KARL-HEINZ SEIFERT. **Researches on the Problem of Compensation in Manual Dexterity**

GOTTFRIED PETRI. **Differential Analysis**

WALTER KRISTOF. **Statistical Method in Judging Profile Tests**

WILHELM ARNOLD. **New Experiments with the Pauli Test**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per issue.

Subscription: DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN

Brentanoweg 10, Germany